



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> : <b>C12N 15/57, 9/48, C12Q 1/68, C12N 5/10, C07K 16/40, A61K 38/48, 38/55</b>		<b>A2</b>	(11) International Publication Number: <b>WO 00/09709</b>
			(43) International Publication Date: 24 February 2000 (24.02.00)
(21) International Application Number: PCT/US99/17818 (22) International Filing Date: 6 August 1999 (06.08.99)  (30) Priority Data: 60/096,114                      10 August 1998 (10.08.98)                      US 60/119,768                      11 February 1999 (11.02.99)                      US  (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications US    60/096,114 (CIP) Filed on                                      10 August 1998 (10.08.98) US    60/119,768 (CIP) Filed on                                      11 February 1999 (11.02.99)  (71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).  (72) Inventors; and (75) Inventors/Applicants (for US only): BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577		(US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). REDDY, Roopa [IN/US]; 1233 W. McKinley Drive, Sunnyvale, CA 94086 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). AU-YOUNG, Janice [US/US]; 1419 Kains Avenue, Berkeley, CA 94702 (US). SHIH, Leo, L. [US/US]; Apartment B., 1081 Tanland Drive, Palo Alto, CA 94303 (US). LU, Dyung, Aina, M. [US/US]; 55 Park Belmont Place, San Jose, CA 95136 (US).  (74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).  (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  Published Without international search report and to be republished upon receipt of that report.	
(54) Title: PROTEASES AND ASSOCIATED PROTEINS			
(57) Abstract			
The invention provides human proteases and associated proteins (PPRG) and polynucleotides which identify and encode PPRG. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of PPRG.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakistan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## PROTEASES AND ASSOCIATED PROTEINS

### TECHNICAL FIELD

5           This invention relates to nucleic acid and amino acid sequences of proteases and associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative and immune disorders.

### BACKGROUND OF THE INVENTION

10           Proteolytic processing is an essential component of normal cell growth, differentiation, remodeling, and homeostasis. The cleavage of peptide bonds within cells is necessary for the maturation of precursor proteins to their active forms, the removal of signal sequences from targeted proteins, the degradation of incorrectly folded proteins, and the controlled turnover of peptides within the cell. Proteases participate in apoptosis, inflammation, and tissue remodeling  
15 during embryonic development, wound healing, and normal growth. They are necessary components of bacterial, parasitic, and viral invasion and replication within a host. Four principal categories of mammalian proteases have been identified based on active site structure, mechanism of action, and overall three-dimensional structure. (See Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp. 1-5.)

20           The serine proteases (SPs) are a large family of proteolytic enzymes that include the digestive enzymes, trypsin and chymotrypsin; components of the complement cascade and of the blood-clotting cascade; and enzymes that control the degradation and turnover of macromolecules of the extracellular matrix. SPs are so named because of the presence of a serine residue in the active site for catalysis of protein cleavage. The active site of an SP is composed of a triad of  
25 residues including the aforementioned serine, an aspartate, and a histidine residue. SPs have a wide range of substrate specificities and can be subdivided into subfamilies on the basis of these specificities. The main sub-families are trypases which cleave after arginine or lysine; aspases which cleave after aspartate; chymases which cleave after phenylalanine or leucine; metases which cleavage after methionine; and serases which cleave after serine. Clp protease is a unique member  
30 of the serine protease family as its activity is controlled by a regulatory subunit that binds and hydrolyzes ATP. Clp protease was originally found in plant chloroplasts but is believed to be widespread in both prokaryotic and eukaryotic cells (Maurizi, M.R. et al. (1990) J. Biol. Chem. 265:12546-12552). SKD3, a mammalian homolog of the bacterial Clp regulatory subunit, has recently been identified in mouse (Perier, F. et al. (1995) Gene 152:157-163).

Cysteine proteases are involved in diverse cellular processes ranging from the processing of precursor proteins to intracellular degradation. Mammalian cysteine proteases include lysosomal cathepsins and cytosolic calcium activated proteases, calpains. Of particular note, cysteine proteases are produced by monocytes, macrophages and other cells of the immune system which migrate to sites of inflammation and, in their protective role, secrete various molecules to repair damaged tissue. These cells may overproduce the same molecules and cause tissue destruction in certain disorders. In autoimmune diseases such as rheumatoid arthritis, the secretion of the cysteine protease, cathepsin C, degrades collagen, laminin, elastin and other structural proteins found in the extracellular matrix of bones. The cathepsin family of lysosomal proteases includes the cysteine proteases: cathepsins B, H, K, L, O2, and S; and the aspartyl proteases: cathepsins D and G. Various members of this endosomal protease family are differentially expressed. Some, such as cathepsin D, have a ubiquitous tissue distribution while others, such as cathepsin L, are found only in monocytes, macrophages, and other cells of the immune system.

Aspartic proteases include bacterial penicillopepsin, mammalian pepsin, renin, chymosin, and certain fungal proteases. The characteristic active site residues of aspartic proteases are a pair of aspartic acid residues, for example, Asp33 and Asp213 in penicillopepsin. Aspartic proteases are also called acid proteases because the optimum pH for their activity is between 2 and 3. In this pH range, one of the aspartate residues is ionized and the other is neutral. A potent inhibitor of aspartic proteases is the hexapeptide pepstatin which, in the transition state, resembles normal substrates.

Carboxypeptidases A and B are the principal mammalian representatives of the metalloprotease family. Both are exopeptidases of similar structure and active site configuration. Carboxypeptidase A, like chymotrypsin, prefers C-terminal aromatic and aliphatic side chains of hydrophobic nature, whereas carboxypeptidase B is directed toward basic arginine and lysine residues. Active site components include zinc, which coordinates one histidine and two glutamic acid residues in the protein.

Proteasomes and ubiquitin proteases are both associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins in eukaryotic cells and some bacteria. Proteasomes are large (~2000 kDa), multisubunit complexes composed of a central catalytic core containing a variety of proteases, and terminal subunits that serve in substrate recognition and regulation of proteasome activity. The UCS mediates the elimination of abnormal proteins and regulates the half-lives of important regulatory proteins that control cellular processes such as gene transcription and cell cycle progression. In the UCS pathway, a protein targeted for



degradation is conjugated to ubiquitin, a small, heat-stable protein. The ubiquitinated protein is then recognized and degraded by a proteasome, and ubiquitin is released by ubiquitin protease for reutilization in the UCS. The UCS is implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes such as p53, viral proteins, cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, A. (1994) *Cell* 79:13-21). A murine proto-oncogene, Unp, encodes a nuclear ubiquitin protease whose overexpression leads to oncogenic transformation of NIH 3T3 cells, and the human homolog of this gene is consistently elevated in small cell tumors and adenocarcinomas of the lung (Gray, D.A. (1995) *Oncogene* 10:2179-2183).

Many other proteolytic enzymes do not fit any of the major categories discussed above because their mechanisms of action and/or active sites have not been elucidated. These include the aminopeptidases and signal peptidases. Aminopeptidases catalyze the hydrolysis of amino acid residues from the amino terminus of peptide substrates. Bovine leucine aminopeptidase is a zinc metalloenzyme that utilizes the sulfhydryl groups from at least three reactive cysteine residues at its active site in the binding of metal ions (Cuypers, H.T. et al. (1982) *J. Biol. Chem.* 257:7086-7091).

Signal peptidases are a specialized class of proteases found in all prokaryotic and eukaryotic cell types that serve in the processing of signal peptides. Signal peptides are amino-terminal sequences which direct the protein from its ribosomal assembly site to a particular cellular or extracellular location. Once the protein has been exported, removal of the signal sequence by a signal peptidase and posttranslational processing activate the protein. Signal peptidases exist as multi-subunit complexes in both yeast and mammals.

Protease inhibitors and other regulators of protease activity control the activity and effects of proteases. Protease inhibitors have been shown to control pathogenesis in animal models of proteolytic disorders (Murphy, G. (1991) *Agents Actions Suppl.* 35:69-76). Low levels of the cystatins, low molecular weight inhibitors of the cysteine proteases, correlate with malignant progression of tumors. (Calkins, C. et al. (1995) *Biol. Biochem. Hoppe Seyler* 376:71-80). Also, increases in cysteine protease levels, when accompanied by reductions in inhibitor activity, are correlated with the pathology of arthritis and immunological diseases in humans.

Serpins are inhibitors of mammalian plasma serine proteases. Many serpins serve to regulate the blood clotting cascade and/or the complement cascade in mammals. Sp32 is a positive regulator of the mammalian acrosomal protease, acrosin. Sp32 binds the proenzyme, proacrosin, and thereby aides in packaging the enzyme into the acrosomal matrix (Baba, T. et al. (1994) *J. Biol. Chem.* 269:10133-10140).

The Kunitz family of serine protease inhibitors is characterized by one or more "Kunitz domains" containing a series of cysteine residues that are regularly spaced over approximately 50 amino acid residues and form three intrachain disulfide bonds. Members of this family include aprotinin, tissue factor pathway inhibitor (TFPI-1 and TFPI-2), inter- $\alpha$ -trypsin inhibitor, and  
5 bikunin (Marlor, C.W. et al. (1997) J. Biol. Chem. 272:12202-12208). Members of this family are potent inhibitors (in the nanomolar range) against serine proteases such as kallikrein and plasmin.

The discovery of new proteases and associated proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative and immune disorders.

10

### SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, proteases and associated proteins referred to collectively as "PPRG" and individually as "PPRG-1," "PPRG-2," "PPRG-3," "PPRG-4," "PPRG-5," "PPRG-6," "PPRG-7," "PPRG-8," "PPRG-9," "PPRG-10," "PPRG-11,"  
15 "PPRG-12," "PPRG-13," "PPRG-14," "PPRG-15," "PPRG-16," "PPRG-17," "PPRG-18," "PPRG-19," and "PPRG-20." In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and fragments thereof.

The invention further provides a substantially purified variant having at least 90% amino  
20 acid identity to at least one of the amino acid sequences selected from the group consisting of SEQ ID NO:1-20 and fragments thereof. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity  
25 to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 and fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 and fragments  
30 thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1-20 and fragments thereof.

The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of (a) hybridizing the complement of the

polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

5       The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40 and fragments thereof. The invention also provides an  
10 isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40 and fragments thereof.

      The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the  
15 group consisting of SEQ ID NO:1-20 and fragments thereof. In another aspect, the expression vector is contained within a host cell.

      The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide under conditions suitable for the expression of the polypeptide; and  
20 (b) recovering the polypeptide from the host cell culture.

      The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-20 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

      The invention further includes a purified antibody which binds to a polypeptide selected  
25 from the group consisting of SEQ ID NO:1-20 and fragments thereof. The invention also provides a purified agonist and a purified antagonist to the polypeptide.

      The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of PPRG, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a  
30 substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-20 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

      The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of PPRG, the method comprising administering to a subject in

need of such treatment an effective amount of an antagonist of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 and fragments thereof.

### BRIEF DESCRIPTION OF THE TABLES

5 Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding PPRG.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods and algorithms used for identification of PPRG.

10 Table 3 shows the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis, diseases, disorders, or conditions associated with these tissues, and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding PPRG were isolated.

15 Table 5 shows the tools, programs, and algorithms used to analyze PPRG, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is  
20 understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a,"  
25 "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same  
30 meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the  
35 publications and which might be used in connection with the invention. Nothing herein is to be

construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

“PPRG” refers to the amino acid sequences of substantially purified PPRG obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term “agonist” refers to a molecule which, when bound to PPRG, increases or prolongs the duration of the effect of PPRG. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of PPRG.

An “allelic variant” is an alternative form of the gene encoding PPRG. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding PPRG include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same as PPRG or a polypeptide with at least one functional characteristic of PPRG. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PPRG, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PPRG. The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PPRG. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PPRG is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms “amino acid” and “amino acid sequence” refer to an oligopeptide, peptide,

polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of PPRG which are preferably at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity or immunological activity of PPRG. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which, when bound to PPRG, decreases the amount or the duration of the effect of the biological or immunological activity of PPRG. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of PPRG.

The term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind PPRG polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to

the antisense strand, and the designation "positive" can refer to the sense strand.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic PPRG, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PPRG or fragments of PPRG may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

The term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding PPRG, by northern analysis is indicative of the presence of nucleic acids encoding PPRG in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding PPRG.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" and "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR, Madison WI) which creates alignments between two or more sequences according to methods selected by the user, e.g., the clustal method. (See, e.g., Higgins, D.G. and P.M. Sharp (1988) Gene 73:237-244.) Parameters for each method may be the default parameters provided by MEGALIGN or may be specified by the user. The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues



in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) Methods Enzymol. 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term “humanized antibody” refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization” refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term “hybridization complex” refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words “insertion” and “addition” refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

“Immune response” can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term “microarray” refers to an arrangement of distinct polynucleotides on a substrate.

The terms “element” and “array element” in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term “modulate” refers to a change in the activity of PPRG. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PPRG.

The phrases “nucleic acid” or “nucleic acid sequence,” as used herein, refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, “fragments” refers to those nucleic acid sequences which comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:21-40, for example, as distinct from any other sequence in the same genome. For example, a fragment of SEQ ID NO:21-40 is useful in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:21-40 from related polynucleotide sequences. A fragment of SEQ ID NO:21-40 is at least about 15-20 nucleotides in length. The precise length of the fragment of SEQ ID NO:21-40 and the region of SEQ ID NO:21-40 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment. In some cases, a fragment, when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

The terms “operably associated” and “operably linked” refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the translation of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term “oligonucleotide” refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. “Oligonucleotide” is substantially equivalent to the terms “amplimer,” “primer,” “oligomer,” and “probe,” as these terms are commonly defined in the art.

“Peptide nucleic acid” (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

The term “sample” is used in its broadest sense. A sample suspected of containing nucleic acids encoding PPRG, or fragments thereof, or PPRG itself, may comprise a bodily fluid; an

extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

10 The term "stringent conditions" refers to conditions which permit hybridization between polynucleotides and the claimed polynucleotides. Stringent conditions can be defined by salt concentration, the concentration of organic solvent, e.g., formamide, temperature, and other conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

20 A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for

limited periods of time.

A "variant" of PPRG polypeptides refers to an amino acid sequence that is altered by one or more amino acid residues. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with  
5 isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

10 The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to PPRG. This definition may also include, for example, "allelic" (as defined above), "splice," "species," or "polymorphic" variants. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The  
15 corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide  
20 polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

## THE INVENTION

The invention is based on the discovery of new human proteases and associated proteins  
25 (PPRG), the polynucleotides encoding PPRG, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative and immune disorders.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding PPRG. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte  
30 clones in which nucleic acids encoding each PPRG were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. The clones in column 5 were used to assemble the consensus nucleotide sequence of each PPRG and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows the identity of each polypeptide; and column 7 shows analytical methods used to identify each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding PPRG. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists tissue categories which express PPRG as a fraction of total tissue categories expressing PPRG. Column 3 lists diseases, disorders, or conditions associated with those tissues expressing PPRG. Column 4 lists the vectors used to subclone the cDNA library. Of particular note is the kidney-specific expression of SEQ ID NO:29 in 5 out of 7 libraries (71%). Also of note is expression of SEQ ID NO:34 in cervical tumor libraries (60%).

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding PPRG were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

The following fragments of the nucleotide sequences encoding PPRG are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:21-40 and to distinguish between SEQ ID NO:21-40 and related polynucleotide sequences. The useful fragments include the fragment of SEQ ID NO:21 from about nucleotide 1 to about nucleotide 56; the fragment of SEQ ID NO:22 from about nucleotide 161 to about nucleotide 213; the fragment of SEQ ID NO:23 from about nucleotide 110 to about nucleotide 158; the fragment of SEQ ID NO:24 from about nucleotide 117 to about nucleotide 174; the fragment of SEQ ID NO:25 from about nucleotide 191 to about nucleotide 245; the fragment of SEQ ID NO:26 from about nucleotide 204 to about nucleotide 269; the fragment of SEQ ID NO:27 from about nucleotide 181 to about nucleotide 221; the fragments of SEQ ID NO:28 from about nucleotide 509 to about nucleotide 553, and from about nucleotide 1751 to about nucleotide 1795; the fragment of SEQ ID NO:29 from about nucleotide 326 to about nucleotide 370; the fragment of SEQ ID NO:30 from about nucleotide 1190 to about nucleotide 1234; the fragment of SEQ ID NO:31 from about nucleotide 283 to about nucleotide 324; the fragment of SEQ ID NO:32 from about nucleotide 380 to about nucleotide 424; the fragments of SEQ ID NO:33 from about nucleotide 272 to about

nucleotide 316, and from about nucleotide 1784 to about nucleotide 1831; the fragment of SEQ ID NO:34 from about nucleotide 217 to about nucleotide 261; the fragment of SEQ ID NO:35 from about nucleotide 2397 to about nucleotide 2441; the fragment of SEQ ID NO:36 from about nucleotide 218 to about nucleotide 262; the fragments of SEQ ID NO:37 from about nucleotide 165 to about nucleotide 209, and from about nucleotide 651 to about nucleotide 695; the fragment of SEQ ID NO:38 from about nucleotide 812 to about nucleotide 856; the fragment of SEQ ID NO:39 from about nucleotide 541 to about nucleotide 585; and the fragment of SEQ ID NO:40 from about nucleotide 163 to about nucleotide 207. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides.

10       The invention also encompasses PPRG variants. A preferred PPRG variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the PPRG amino acid sequence, and which contains at least one functional or structural characteristic of PPRG.

      The invention also encompasses polynucleotides which encode PPRG. In a particular  
15   embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:21-40, which encodes PPRG.

      The invention also encompasses a variant of a polynucleotide sequence encoding PPRG. In particular, such a variant polynucleotide sequence will have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the  
20   polynucleotide sequence encoding PPRG. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:21-40 which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:21-40. Any one of the polynucleotide variants described  
25   above can encode an amino acid sequence which contains at least one functional or structural characteristic of PPRG.

      It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PPRG, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be  
30   produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PPRG, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PPRG and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring PPRG under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PPRG or its derivatives possessing a substantially different codon usage, e.g.,  
5 inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PPRG and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more  
10 desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PPRG and PPRG derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell  
15 systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PPRG or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:21-40 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M.  
20 and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while  
25 high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion  
30 of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a

most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50 % formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash  
5 stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include  
10 temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS.  
15 Additional variations on these conditions will be readily apparent to those skilled in the art.

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or  
20 combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the Hamilton MICROLAB 2200 (Hamilton, Reno NV), Peltier thermal cycler 200 (PTC200; MJ Research, Watertown MA) and the ABI CATALYST 800 (Perkin-Elmer). Sequencing is then carried out using either ABI 373 or 377  
25 DNA sequencing systems (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY,  
30 pp. 856-853.)

The nucleic acid sequences encoding PPRG may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence



from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR.

10 Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available

15 software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include

20 sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular,

25 capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display

30 may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PPRG may be cloned in recombinant DNA molecules that direct expression of PPRG, or fragments or functional equivalents thereof, in appropriate host cells. Due to the

inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PPRG.

The nucleotide sequences of the present invention can be engineered using methods  
5 generally known in the art in order to alter PPRG-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create  
10 new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding PPRG may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.)  
15 Alternatively, PPRG itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of PPRG, or any part thereof, may be altered during direct synthesis and/or combined with sequences  
20 from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH  
25 Freeman, New York NY.)

In order to express a biologically active PPRG, the nucleotide sequences encoding PPRG or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers,  
30 constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PPRG. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PPRG. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PPRG and its initiation

codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous  
5 translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct  
10 expression vectors containing sequences encoding PPRG and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons,  
15 New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PPRG. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression  
20 vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected  
25 depending upon the use intended for polynucleotide sequences encoding PPRG. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PPRG can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or pSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PPRG into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure  
30 for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of PPRG are needed, e.g. for the production of antibodies, vectors which direct high level expression of PPRG

may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PPRG. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH  
5 promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Grant et al. (1987) *Methods Enzymol.* 153:516-54; and Scorer, C. A. et al. (1994) *Bio/Technology* 12:181-184.)

10 Plant systems may also be used for expression of PPRG. Transcription of sequences encoding PPRG may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680;  
15 Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases  
20 where an adenovirus is used as an expression vector, sequences encoding PPRG may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PPRG in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81:3655-3659.) In addition, transcription enhancers, such as  
25 the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb  
are constructed and delivered via conventional delivery methods (liposomes, polycationic amino  
30 polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PPRG in cell lines is preferred. For example, sequences encoding PPRG can be transformed into cell lines using expression vectors which may contain viral origins of replication

and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* or *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. U.S.A. 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. U.S.A. 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PPRG is inserted within a marker gene sequence, transformed cells containing sequences encoding PPRG can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PPRG under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PPRG and that express PPRG may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or

protein sequences.

Immunological methods for detecting and measuring the expression of PPRG using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and  
5 fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PPRG is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub.  
10 Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PPRG  
15 include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PPRG, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures  
20 may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

25 Host cells transformed with nucleotide sequences encoding PPRG may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PPRG may be designed to contain signal  
30 sequences which direct secretion of PPRG through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro"

form of the protein may also be used to specify protein targeting, folding, and/or activity.

Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Bethesda MD) and may be chosen to ensure the

5 correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PPRG may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PPRG protein containing a heterologous moiety that can be recognized by a commercially available antibody  
10 may facilitate the screening of peptide libraries for inhibitors of PPRG activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification  
15 of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PPRG encoding sequence  
20 and the heterologous protein sequence, so that PPRG may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PPRG may be  
25 achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract systems (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, preferably <sup>35</sup>S-methionine.

Fragments of PPRG may be produced not only by recombinant production, but also by  
30 direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of PPRG may be synthesized separately and then combined to produce the full length molecule.

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PPRG and proteases and associated proteins. In addition, the expression of PPRG is closely associated with cell proliferative conditions, including cancer, and with  
5 inflammation and the immune response. Therefore, PPRG appears to play a role in cell proliferative and immune disorders. In the treatment of cell proliferative and immune disorders associated with increased PPRG expression or activity, it is desirable to decrease the expression or activity of PPRG. In the treatment of the above conditions associated with decreased PPRG expression or activity, it is desirable to increase the expression or activity of PPRG.

10 Therefore, in one embodiment, PPRG or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PPRG. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria,  
15 polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an immune disorder  
20 such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia  
25 with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic  
30 lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma.

In another embodiment, a vector capable of expressing PPRG or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased



expression or activity of PPRG including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified PPRG in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PPRG  
5 including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PPRG may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PPRG including, but not limited to, those listed above.

In a further embodiment, an antagonist of PPRG may be administered to a subject to treat  
10 or prevent a disorder associated with increased expression or activity of PPRG. Examples of such disorders include, but are not limited to, those described above. In one aspect, an antibody which specifically binds PPRG may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express PPRG.

In an additional embodiment, a vector expressing the complement of the polynucleotide  
15 encoding PPRG may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PPRG including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination  
20 therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

25 An antagonist of PPRG may be produced using methods which are generally known in the art. In particular, purified PPRG may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PPRG. Antibodies to PPRG may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments,  
30 and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PPRG or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various

adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PPRG have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of PPRG amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PPRG may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. U.S.A. 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. U.S.A. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PPRG-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. U.S.A. 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PPRG may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by

pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

5           Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PPRG and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal  
10 antibodies reactive to two non-interfering PPRG epitopes is preferred, but a competitive binding assay may also be employed (Pound, supra).

          Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PPRG. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of PPRG-antibody complex  
15 divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PPRG epitopes, represents the average affinity, or avidity, of the antibodies for PPRG. The  $K_a$  determined for a preparation of monoclonal  
20 antibodies, which are monospecific for a particular PPRG epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the PPRG-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PPRG, preferably in active form, from the antibody (Catty, D. (1988) Antibodies,  
25 Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J.E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

          The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml,  
30 preferably 5-10 mg specific antibody/ml, is preferred for use in procedures requiring precipitation of PPRG-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

          In another embodiment of the invention, the polynucleotides encoding PPRG, or any

fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding PPRG may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding PPRG. Thus, complementary molecules  
5 or fragments may be used to modulate PPRG activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PPRG.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses,  
10 or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding PPRG. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding PPRG can be turned off by transforming a cell or tissue with expression  
15 vectors which express high levels of a polynucleotide, or fragment thereof, encoding PPRG. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication  
20 elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding PPRG. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred.  
25 Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be  
30 designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For

example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PPRG.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences:

- 5 GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

- 10 Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PPRG. Such DNA sequences may be  
15 incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

- RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3'  
20 ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by  
25 endogenous endonucleases.

- Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers  
30 may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotech. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of PPRG, antibodies to PPRG, and mimetics, agonists, antagonists, or inhibitors of PPRG. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or

solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol.

- 5 Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

- 10 Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include
- 15 fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

- For topical or nasal administration, penetrants appropriate to the particular barrier to be
- 20 permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

- The pharmaceutical composition may be provided as a salt and can be formed with many
- 25 acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to
- 30 use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of PPRG, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions

wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in  
5 cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example  
10 PPRG or fragments thereof, antibodies of PPRG, and agonists, antagonists or inhibitors of PPRG, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic  
15 effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending  
20 upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and  
25 gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of  
30 about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.



## DIAGNOSTICS

In another embodiment, antibodies which specifically bind PPRG may be used for the diagnosis of disorders characterized by expression of PPRG, or in assays to monitor patients being treated with PPRG or agonists, antagonists, or inhibitors of PPRG. Antibodies useful for  
5 diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PPRG include methods which utilize the antibody and a label to detect PPRG in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known  
10 in the art and may be used.

A variety of protocols for measuring PPRG, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PPRG expression. Normal or standard values for PPRG expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to  
15 PPRG under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of PPRG expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PPRG may be used  
20 for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of PPRG may be correlated with disease. The diagnostic assay may be used to determine absence,  
25 presence, and excess expression of PPRG, and to monitor regulation of PPRG levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PPRG or closely related molecules may be used to identify nucleic acid sequences which encode PPRG. The specificity of  
30 the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding PPRG, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably

have at least 50% sequence identity to any of the PPRG encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:21-40 or from genomic sequences including promoters, enhancers, and introns of the PPRG gene.

- 5 Means for producing specific hybridization probes for DNAs encoding PPRG include the cloning of polynucleotide sequences encoding PPRG or PPRG derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a
- 10 variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

- Polynucleotide sequences encoding PPRG may be used for the diagnosis of disorders associated with expression of PPRG. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis,
- 15 cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas,
- 20 parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an immune disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis,
- 25 contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis,
- 30 pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma. The polynucleotide sequences encoding PPRG may be used in

Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PPRG expression. Such qualitative or quantitative methods are well known in the art.

5           In a particular aspect, the nucleotide sequences encoding PPRG may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PPRG may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated  
10 and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PPRG in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

15           In order to provide a basis for the diagnosis of a disorder associated with expression of PPRG, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PPRG, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from  
20 normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

          Once the presence of a disorder is established and a treatment protocol is initiated,  
25 hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

          With respect to cancer, the presence of an abnormal amount of transcript (either under- or  
30 overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PPRG may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PPRG, or a fragment of a polynucleotide complementary to the  
5 polynucleotide encoding PPRG, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantify the expression of PPRG include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and  
10 interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

15 In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and  
20 to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad.  
25 Sci. U.S.A. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding PPRG may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes  
30 (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical

chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding PPRG on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PPRG, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PPRG and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PPRG, or fragments thereof, and washed. Bound PPRG is then detected by methods well known in the art. Purified PPRG can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PPRG specifically compete with a test compound for

binding PPRG. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PPRG.

In additional embodiments, the nucleotide sequences which encode PPRG may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/096,114 and U.S. Ser. No. 60/119,768, are hereby expressly incorporated by reference.

15

## EXAMPLES

### I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies),

using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA  
5 was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), pSPORT1 plasmid (Life Technologies), or pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids  
10 were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision, using the UNIZAP vector system (Stratagene) or cell lysis. Plasmids were purified using at least one of the following: a  
15 Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

20 Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a Fluoroskan II  
25 fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins  
30 Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA

sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing systems (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the  
5 cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows  
10 the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using  
15 MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR).

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then  
20 queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length  
25 polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families.  
30 (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:21-40. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.



#### IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; 5 Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is 10 categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

100

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact 15 within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding PPRG occurred. Analysis involved the categorization of cDNA 20 libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation/trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of 25 libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

#### V. Extension of PPRG Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:21-27 were produced by extension 30 of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. For each nucleic acid sequence, one primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of

interest. The initial primers were designed from the cDNA using OLIGO 4.06 (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and  
 5 primer-primer dimerizations was avoided.

Selected human cDNA libraries (Life Technologies) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR kit  
 10 (Perkin-Elmer) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the PTC200 thermal cycler (M.J. Research) beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, with the following parameters:

	Step 1	94°C for 1 min (initial denaturation)
	Step 2	65°C for 1 min
15	Step 3	68°C for 6 min
	Step 4	94°C for 15 sec
	Step 5	65°C for 1 min
	Step 6	68°C for 7 min
	Step 7	Repeat steps 4-6 for an additional 15 cycles
20	Step 8	94°C for 15 sec
	Step 9	65°C for 1 min
	Step 10	68°C for 7:15 min
	Step 11	Repeat steps 8-10 for an additional 12 cycles
	Step 12	72°C for 8 min
25	Step 13	4°C (and holding)

A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the  
 30 gel, purified using the QIAQUICK kit (QIAGEN), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13  $\mu$ l of ligation buffer, 1  $\mu$ l T4-DNA ligase (15 units) and 1  $\mu$ l T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16°C. Competent *E. coli* cells (in  
 35 40  $\mu$ l of appropriate media) were transformed with 3  $\mu$ l of ligation mixture and cultured in 80  $\mu$ l of SOC medium. (See, e.g., Sambrook, *supra*, Appendix A, p. 2.) After incubation for one hour at 37°C, the *E. coli* mixture was plated on Luria Bertani (LB) agar (See, e.g., Sambrook, *supra*, Appendix A, p. 1) containing carbenicillin (2x carb). The following day, several colonies were randomly picked from each plate and cultured in 150  $\mu$ l of liquid LB/2x carb medium placed in an

individual well of an appropriate commercially-available sterile 96-well microtiter plate. The following day, 5  $\mu$ l of each overnight culture was transferred into a non-sterile 96-well plate and, after dilution 1:10 with water, 5  $\mu$ l from each sample was transferred into a PCR array.

For PCR amplification, 18  $\mu$ l of concentrated PCR reaction mix (3.3x) containing 4 units  
 5 of rTth DNA polymerase, a vector primer, and one or both of the gene-specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

	Step 1	94°C for 60 sec
	Step 2	94°C for 20 sec
10	Step 3	55°C for 30 sec
	Step 4	72°C for 90 sec
	Step 5	Repeat steps 2-4 for an additional 29 cycles
	Step 6	72°C for 180 sec
15	Step 7	4°C (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequence of SEQ ID NO:21-27 are used to obtain 5'  
 20 regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

The full length nucleic acid sequences of SEQ ID NO:28-40 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment,  
 25 and the other primer to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

30 Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  
 35  $(NH_4)_2SO_4$ , and  $\beta$ -mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec;

Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

5       The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture  
10       was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

      The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and  
15       sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in  
20       restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

      The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following  
25       parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulphoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer  
30       sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

      In like manner, the nucleotide sequences of SEQ ID NO:28-40 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

## VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:21-40 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [5 5 10  $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography and compared.

## VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g.,

Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

### VIII. Complementary Polynucleotides

- 5 Sequences complementary to the PPRG-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PPRG. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of
- 10 PPRG. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PPRG-encoding transcript.

### IX. Expression of PPRG

- 15 Expression and purification of PPRG is achieved using bacterial or virus-based expression systems. For expression of PPRG in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac*
- 20 operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PPRG upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PPRG in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of
- 25 baculovirus is replaced with cDNA encoding PPRG by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic
- 30 modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. U.S.A. 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PPRG is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates.

GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PPRG at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch 10 and 16). Purified PPRG obtained by these methods can be used directly in the following activity assay.

10 **X. Demonstration of PPRG Activity**

Protease activity of PPRG is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore. (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp.25-55.) Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (Carboxypeptidase A and B, procollagen C-proteinase). Chromogens commonly used are 2-naphthylamine, 4-nitroaniline, and furylacrylic acid. Assays are performed at ambient temperature and contain an aliquot of the enzyme and the appropriate substrate in a suitable buffer. Reactions are carried out in an optical cuvette and followed by measurement of the increase/decrease in absorbance of the chromogen released during hydrolysis of the peptide substrate. The change in absorbance is proportional to the enzyme activity in the assay.

Regulation of protease activity (agonism or antagonism) by PPRG is measured using an appropriate protease assay as described above in the presence or absence of PPRG as an agonist or inhibitor of this activity. Protease activity is measured in the absence of PPRG (control activity) and in the presence of varying amounts of PPRG. The change in protease activity compared to the control is proportional to the amount of PPRG in the assay and is a measure of the protease regulatory activity of PPRG.

30 For example, for inhibitory activity of PPRG-2, the assay is carried out as described above for PPRG using a calcium activated protease, such as calpain, assayed in the absence and in the presence of various concentrations of PPRG-2. Inhibition of calpain protease activity is proportional to the activity of PPRG-2 in the assay. Similarly, for inhibitory activity of PPRG-4 and PPRG-9, assays are carried out as described above for PPRG using pancreatic trypsin assayed

in the absence and in the presence of various concentrations of PPRG-4 or PPRG-9. Inhibition of pancreatic trypsin protease activity is proportional to the activity of PPRG-4 or PPRG-9 in the assay.

#### **XI. Functional Assays**

- 5 PPRG function is assessed by expressing the sequences encoding PPRG at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of
- 10 recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of
- 15 choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear
- 20 DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of
- 25 fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PPRG on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PPRG and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved

30 regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PPRG and other genes of interest can be analyzed by northern analysis or microarray techniques.



## **XII. Production of PPRG Specific Antibodies**

PPRG substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

5 Alternatively, the PPRG amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

10 Typically, oligopeptides 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for anti-peptide activity  
15 by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

## **XIII. Purification of Naturally Occurring PPRG Using Specific Antibodies**

Naturally occurring or recombinant PPRG is substantially purified by immunoaffinity chromatography using antibodies specific for PPRG. An immunoaffinity column is constructed by  
20 covalently coupling anti-PPRG antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PPRG are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PPRG (e.g., high ionic strength  
25 buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PPRG binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PPRG is collected.

## **XIV. Identification of Molecules Which Interact with PPRG**

PPRG, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter  
30 reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PPRG, washed, and any wells with labeled PPRG complex are assayed. Data obtained using different concentrations of PPRG are used to calculate values for the number, affinity, and association of PPRG with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited  
5 to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	21	1220330	NEUTGMT01	1220330H1 (NEUTGMT01), 1220330R6 (NEUTGMT01), 3031706F6 (TLYMNOT05)
2	22	1342493	COLNTUT03	071068F1 (PLACNOB01), 1321108F6 (BLADNOT04), 1342493F6 (COLNTUT03), 1342493H1 (COLNTUT03), 1345967T6 (PROSNOT15), 1438889F1 (PANCNOT08), 1679890T7 (STOMFET01), 1800338T6 (COLNNOT27), 3217273H1 (TESTNOT07)
3	23	1698270	BLADTUT05	1698270H1 (BLADTUT05), 1374869H1 (LUNGNOT10), 312647R1 (LUNGNOT02), 386032H1 (THYMNOT02)
4	24	2012492	TESTNOT03	2004918R6 (TESTNOT03), 2004918T6 (TESTNOT03), 2011777H1 (TESTNOT03), 2012492H1 (TESTNOT03)
5	25	2309875	NGANNOT01	1597268F6 (BRAINOT14), 1682605X22C1 (PROSNOT15), 1683253X19C1 (PROSNOT15), 1685583X13C1 (PROSNOT15), 1752982H1 (LIVRTUT01), 2052808F6 (LIVRFET02), 2197089H1 (SPLNFET02), 856284R1 (NGANNOT01), 2309875H1 (NGANNOT01)
6	26	2479394	SMCANOT01	2479394F6 (SMCANOT01), 2479394H1 (SMCANOT01), 2623972X42F1 (KERANOT02), SAEC10649F1, SAEA03168R1, SAEC11168F1, SAEA00641R1, SAEC10266F1, SAEC11328F1
7	27	2613215	SINIUCT01	231698R1 (SINTNOT02), 1363780F1 (LUNGNOT12), 1546635R6 (PROSTUT04), 1662163F6 (BRSTNOT09), 1859908F6 (PROSNOT18), 2192713X13F1 (THYRTUT03), 2192713X15F1 (THYRTUT03), 2543078X303F1 (UTRSNOT11), 2613215H1 (SINIUCT01)
8	28	001528	U937NOT01	001528F1 (U937NOT01), 001528H1 (U937NOT01), 001528X5 (U937NOT01), 001528X6 (U937NOT01), 001528X9 (U937NOT01), 443686R6 (MPHGNOT03), 2596612H1 (OVARTUT02), 2888384X12F1 (LUNGFEF04), 3598232H1 (FIBPNOT01), 4906930H2 (TLYMNOT08)
9	29	998626	KIDNTUT01	998626H1 (KIDNTUT01), 998626R6 (KIDNTUT01), 4073122F6 (KIDNNOT26)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
10	30	1393301	THYRNOT03	1393301H1 (THYRNOT03), 2008519T6 (TESTNOT03), SBFA01183F1, SBFA01807F1, SBFA03248F1, SBFA00528F1
11	31	1444055	THYRNOT03	1444055H1 (THYRNOT03), 1444055R1 (THYRNOT03), 2738343H1 (OVARNOT09)
12	32	1650177	PROSTUT09	1616250F6 (BRAITUT12), 1616250T6 (BRAITUT12), 1650177F6 (PROSTUT09), 1650177H1 (PROSTUT09), 2372255H1 (ADRENOT07), 3286138F6 (HEAONOT05), 4012302H1 (MUSCNOT10), SAEA00123F1
13	33	1902576	OVARNOT07	1902576H1 (OVARNOT07), 2909961H1 (KIDNTUT15), SZAP00669V1, SZAP02354V1, SZAP00959V1, SZAP01377V1, SZAP00432V1, SZAP00726V1, SZAP01982V1
14	34	2024210	KERANOT02	2024210H1 (KERANOT02), 4569479H1 (HELATXT01), 4817326H1 (HELATXT03)
15	35	2523109	BRAITUT21	2523109H1 (BRAITUT21), 3574330H1 (BRONNOT01), 3126142H1 (LUNGNOT12), 3417837H2 (PTHYNOT04), 2309843X13C1 (NGANOT01), 2365785X305D1 (ADRENOT07), 2674631F6 (KIDNNOT19), 4770421H1 (BRATNOT02), 2122564F6 (BRSTNOT07), 5401752H1 (BRAHNOT01), 2196601F6 (SPLNFET02), 2599102F6 (UTRSNOT10), 3030634T6 (HEARFET02), 1721515T6 (BLADNOT06), 546753F1 (BEPINOT01)
16	36	2588566	LUNGNOT22	2588566H1 (LUNGNOT22), 2588566X303D1 (LUNGNOT22), 2727313T6 (OVARNOT05), 3972055H1 (ADRETUT06), SBKA00529F1

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
17	37	2740570	BRSTTUT14	102671F1 (ADRENOR01), 102671R1 (ADRENOR01), 678618X16 (UTRSNOT02), 1259309F6 (MENITUT03), 1466058F6 (PANCTUT02), 2740570H1 (BRSTTUT14), 2740570X316D2 (BRSTTUT14), 2740570X319F1 (BRSTTUT14), 3050368H1 (LUNGNOT25), SCJA02363V1
18	38	2820384	BRSTNOT14	1642163F6 (HEARFET01), 1706505F6 (DUODNOT02), 1742853T6 (HIPONON01), 1853454F6 (LUNGFET03), 1878661F6 (LEUKNOT03), 1878661H1 (LEUKNOT03), 2820384H1 (BRSTNOT14), 2820384X13F1 (BRSTNOT14), 3497393H1 (PROSTUT13), 3633187H1 (LIVRNOT03), 4059719H1 (BRAINOT21), 4144331H1 (BRSTTMT01), 4982538H1 (HELATXT05)
19	39	2990692	KIDNFET02	2990692F6 (KIDNFET02), 2990692H1 (KIDNFET02), 2990692X14F1 (KIDNFET02), 2990692X34F1 (KIDNFET02), 4636147H1 (MYEPTXT01)
20	40	4590384	MASTTXT01	1487107F6 (UCMCL5T01), 4590384H1 (MASTTXT01), 4918570H1 (LIVRFET05), SANA01269F1

Table 2

Polypeptide Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
1	206	T66 S38 T103 T154 S180 T21 T31 T68 T84			Metalloproteinase	BLAST BLOCKS PRINTS
2	754	S29 T79 S188 S197 T216 T224 T235 T331 S357 T391 S410 T474 S607 S609 S709 T717 T744 S13 S42 T63 S87 S139 S167 S194 S268 T297 T313 T435 T470 S728 S741 T748 S573 T681 T687	N220 N570		Calpastatin	BLAST
3	308	S33 S136 S207 T220 S290 S304 S41 T122 S125 Y268	N144 N167	Prolyl aminopeptidase: L105 Serine protease: L66	Protease	BLOCKS PRINTS
4	164	T157		Kunitz family signature: F136	Trypsin inhibitor	BLAST BLOCKS MOTIFS PRINTS
5	565	T155 T451 S477 S115 S298 S350 T392 S415 T424 S488 T150 S156 S171 S187 S232 S415 S446 T447 S472 S494 Y195	N509 N533	Ubiquitin carboxyl- terminal hydrolase family 2 signature: Y502	Ubiquitin specific protease 41	BLAST BLOCKS MOTIFS PRINTS

Table 2 (cont.)

Polypeptide Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
6	421	T90 S210 S284 S290 S346 S365 T401 T411 T165 T194 S321 Y310	N260	Zinc carboxy- peptidase, zinc- binding region signatures: P172, H308	Carboxypeptidase	BLAST BLOCKS MOTIFS PRINTS
7	666	T36 S97 T145 S220 T243 S257 S289 S326 S404 S450 T480 S522 T551 S619 T621 T634 S4 T199 S334 T445 S548	N132 N446		Aminopeptidase P	BLAST BLOCKS MOTIFS PRINTS
8	952	S153 S810 T105 S170 T197 S312 S513 T593 S623 S625 S636 S644 S649 T767 T821 T885 S932 T11 S23 S78 T149 S322 T329 T670 T790 Y31 Y578 Y779 Y876		Ubiquitin hydrolase: G261-L278, Y846-V883	Ubiquitin protease	BLAST MOTIFS PFAM
9	166	S48 S119		Signal peptide: M1-A26	Trypsin inhibitor	BLAST MOTIFS SPSCAN
10	543	S505 S39 T41 S98 T134 T158 T250 S291 S331 S359 S466 T53 T59 T160 T342 S379 S399 S425 S489 Y481		Signal peptide: M1-A25	sp32 precursor, proacrosin- binding protein	BLAST MOTIFS SPSCAN

Table 2 (cont.)

Polypeptide Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
11	83	S18 S6 S22 S40		Caspase: D15-P81	Cysteine protease	BLAST MOTIFS PFAM
12	648	S41 S132 T176 T190 T222 T242 T593 T25 S33 S64 S204 T335 T381 S472 S562 T589 S597 T630 Y263 Y310 Y508		ATP/GTP-binding site: G322-T329 Ankyrin repeat: K206-E238 Chaperonins ClpA/B: L138-I592	SKD3, regulator of Clp protease activity	BLAST MOTIFS PFAM
13	672	S99 T123 S282 S547 S568 T644 T42 T52 T110 T207 S226 T332 T488 S522 T622		Cysteine protease: Q67-A78 Calpain: L13-T322	Calcium (cysteine) protease	BLAST BLOCKS MOTIFS PFAM PRINTS
14	80	S73		Kazal-type serine protease inhibitor: C30-C80	Protease inhibitor	BLAST MOTIFS PFAM PROFILESCAN
15	795	S418 T419 T655 S166 T278 T296 S307 S425 T427 T481 S517 S564 S639 S675 T103 S244 S330 T455 S495 S506 T556 Y138		ATP-dependent Clp protease: A345-A363 Signal peptide: M1-W23 Transmembrane domain: A254-F272	Paraplegin (metalloprotease)	BLAST MOTIFS PFAM PRINTS SPSCAN



Table 2 (cont.)

Polypeptide Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
16	193	S19 S63 T182 S4 T140 T168			Neutral protease alpha subunit	BLAST MOTIFS
17	663	S437 S448 T547 T23 T27 S33 S35 S46 S98 S108 T222 S253 T289 S414 S436 T473 S481 S48 T120 S182 S347		Ubiquitin carboxyl- terminal hydrolase: Y378-V415	Ubiquitin specific protease UBP 41	BLAST MOTIFS PFAM
18	362	S130 T69 S129 T166 S40 S348 Y39		Ubiquitin carboxyl- terminal hydrolase: Y71-V108	Ubiquitin specific protease UBP 41	BLAST MOTIFS PFAM
19	210	T133 T144 T89 S199		Retroviral aspartyl protease: V111-I193	Human endogenous retroviral protease	BLAST MOTIFS PFAM
20	283	S266 S77 S94 T110 S166 S50 S191 S208 T275		Trypsin: I34-I258 Serine protease, active site: V70-C75	Metase (serine protease)	BLAST MOTIFS PFAM PROFILESCAN

Table 3

Nucleotide SEQ ID NO:	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
21	Hematopoietic/Immune (0.750) Reproductive (0.250)	Inflammation (0.750) Cancer (0.250) Fetal (0.250)	PSPORT
22	Reproductive (0.255) Gastrointestinal (0.196) Cardiovascular (0.125)	Cancer (0.475) Inflammation (0.245) Fetal (0.152)	pINCY
23	Reproductive (0.258) Cardiovascular (0.129) Gastrointestinal (0.129)	Cancer (0.419) Inflammation (0.226) Fetal (0.204)	pINCY
24	Reproductive (1.00)	Inflammation (1.000)	PBLUESCRIPT
25	Reproductive (0.258) Nervous (0.210) Gastrointestinal (0.161)	Cancer (0.548) Inflammation (0.242) Fetal (0.129)	PSPORT
26	Nervous (0.500) Cardiovascular (0.250) Dermatologic (0.250)	Cancer (0.500) Fetal (0.500)	pINCY
27	Reproductive (0.244) Gastrointestinal (0.179) Developmental (0.141)	Cancer (0.418) Fetal (0.231) Inflammation (0.154)	pINCY
28	Hematopoietic/Immune (0.304) Reproductive (0.232) Cardiovascular (0.107)	Cell proliferation (0.465) Inflammation (0.429)	PBLUESCRIPT
29	Urologic (0.714) Musculoskeletal (0.147)	Cancer (0.857) Inflammation (0.143)	PSPORT1

Table 3 (cont.)

Nucleotide SEQ ID NO:	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
30	Reproductive (0.375) Endocrine (0.125) Hematopoietic/Immune (0.250)	Inflammation (0.500) Cancer (0.375)	pINCY
31	Hematopoietic/Immune (0.308) Cardiovascular (0.154) Reproductive (0.154)	Inflammation (0.616) Cancer (0.385)	pINCY
32	Hematopoietic/Immune (0.261) Musculoskeletal (0.217) Reproductive (0.217)	Cell proliferation (0.565) Inflammation (0.435)	pINCY
33	Reproductive (0.333) Nervous (0.222)	Cell proliferation (0.703) Inflammation (0.148)	pINCY
34	Reproductive (0.600) Dermatologic (0.300) Nervous (0.100)	Cell proliferation (0.800)	PSPORT1
35	Reproductive (0.202) Nervous (0.173) Gastrointestinal (0.135)	Cell proliferation (0.586) Inflammation (0.279)	pINCY
36	Gastrointestinal (0.500) Cardiovascular (0.333) Endocrine (0.167)	Cancer (0.833) Inflammation (0.167)	pINCY
37	Nervous (0.205) Reproductive (0.205) Cardiovascular (0.179)	Cell proliferation (0.538) Inflammation (0.154)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
38	Hematopoietic/Immune (0.267) Reproductive (0.250) Nervous (0.133)	Cell proliferation (0.600) Inflammation (0.383)	pINCY
39	Hematopoietic/Immune (0.400) Developmental (0.200) Gastrointestinal (0.200)	Cell proliferation (0.800) Inflammation (0.400)	pINCY
40	Gastrointestinal (0.500) Hematopoietic/Immune (0.500)	Cell proliferation (0.500) Inflammation (0.500)	pINCY

Table 4

Nucleotide SEQ ID NO:	Library	Library Comment
21	NEUTGMT01	Library was constructed using RNA isolated from peripheral blood granulocytes collected by density gradient centrifugation through Ficoll-Hypaque. The cells were isolated from buffy coat units obtained from 20 unrelated male and female donors. Cells were cultured in 10 nM GM-CSF for 1 hour before washing and harvesting for RNA preparation.
22	COLNTUT03	Library was constructed using RNA isolated from colon tumor tissue obtained from the sigmoid colon of a 62-year-old Caucasian male during a sigmoidectomy and permanent colostomy. Pathology indicated invasive grade 2 adenocarcinoma. One lymph node contained metastasis with extranodal extension. Patient history included hyperlipidemia, cataract disorder, and dermatitis. Family history included cardiovascular disease and cancer.
23	BLADTUT05	Library was constructed using RNA isolated from bladder tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma on the anterior wall of the bladder. Patient history included lung neoplasm and tobacco abuse in remission. Family history included a malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease, and lung cancer.
24	TESTNOT03	Library was constructed using polyA RNA isolated from testicular tissue removed from a 37-year-old Caucasian male who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
25	NGANNOT01	Library was constructed using RNA isolated from tumorous neuroganglion tissue removed from a 9-year-old Caucasian male during a soft tissue excision of the chest wall. Pathology indicated a ganglioneuroma. Family history included asthma.
26	SMCANOT01	Library was constructed using RNA isolated from an aortic smooth muscle cell line derived from the explanted heart of a male during a heart transplant.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Comment
27	SINIUCT01	Library was constructed using RNA isolated from ileum tissue obtained from a 42-year-old Caucasian male during a total intra-abdominal colectomy and endoscopic jejunostomy. Previous surgeries included polypectomy, colonoscopy, and spinal canal exploration. Family history included cerebrovascular disease, benign hypertension, atherosclerotic coronary artery disease, and type II diabetes.
28	U937NOT01	Library was constructed at Stratagene (STR937207), using RNA isolated from the U937 monocyte-like cell line. This line (ATCC CRL1593) was established from malignant cells obtained from the pleural effusion of a 37-year-old Caucasian male with diffuse histiocytic lymphoma.
29	KIDNTUT01	Library was constructed using RNA isolated from kidney tumor tissue removed from an 8-month-old female during nephroureterectomy. Pathology indicated Wilms' tumor (nephroblastoma), which involved 90 percent of the renal parenchyma. Prior to surgery, the patient was receiving heparin anticoagulant therapy.
30	THYRNOT03	Library was constructed using RNA isolated from thyroid tissue removed from the left thyroid of a 28-year-old Caucasian female during a complete thyroidectomy. Pathology indicated a small nodule of adenomatous hyperplasia present in the left thyroid. Pathology for the associated tumor tissue indicated dominant follicular adenoma forming a well-encapsulated mass in the left thyroid.
31	THYRNOT03	Library was constructed using RNA isolated from thyroid tissue removed from the left thyroid of a 28-year-old Caucasian female during a complete thyroidectomy. Pathology indicated a small nodule of adenomatous hyperplasia present in the left thyroid. Pathology for the associated tumor tissue indicated dominant follicular adenoma forming a well-encapsulated mass in the left thyroid.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Comment
32	PROSTUT09	Library was constructed using RNA isolated from prostate tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma. The patient presented with prostatic inflammatory disease. Patient history included lung neoplasm, and benign hypertension. Family history included a malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease and lung cancer.
33	OVARNOT07	Library was constructed using RNA isolated from left ovarian tissue removed from a 28-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. The tissue was associated with multiple follicular cysts, endometrium in a weakly proliferative phase, and chronic cervicitis of the cervix with squamous metaplasia. Family history included benign hypertension, hyperlipidemia, and atherosclerotic coronary artery disease.
34	KERANOT02	Library was constructed using RNA isolated from epidermal breast keratinocytes (NHEK). NHEK (Clontech #CC-2501) is human breast keratinocyte cell line derived from a 30-year-old black female during breast-reduction surgery.
35	BRAITUT21	Library was constructed using RNA isolated from brain tumor tissue removed from the midline frontal lobe of a 61-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated subfrontal meningotheelial meningioma with no atypia. One ethmoid and mucosal tissue sample indicated meningioma. Family history included cerebrovascular disease, senile dementia, hyperlipidemia, benign hypertension, atherosclerotic coronary artery disease, congestive heart failure, and breast cancer.
36	LUNGNOT22	Library was constructed using RNA isolated from lung tissue removed from a 58-year-old Caucasian female. The tissue sample used to construct this library was found to have tumor contaminant upon microscopic examination. Pathology for the associated tumor tissue indicated a caseating granuloma. Family history included congestive heart failure, breast cancer, secondary bone cancer, acute myocardial infarction and atherosclerotic coronary artery disease.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Comment
37	BRSTTUT14	Library was constructed using RNA isolated from breast tumor tissue removed from a 62-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated an invasive grade 3 (of 4), nuclear grade 3 (of 3) adenocarcinoma, ductal type. Ductal carcinoma in situ, comedo type, comprised 60% of the tumor mass. Metastatic adenocarcinoma was identified in one (of 14) axillary lymph nodes with no perinodal extension. Tumor cells were strongly positive for estrogen receptors and weakly positive for progesterone receptors. Patient history included benign colon neoplasm, hyperlipidemia, cardiac dysrhythmia, and obesity. Family history included atherosclerotic coronary artery disease, myocardial infarction, colon cancer, ovarian cancer, lung cancer, and cerebrovascular disease.
38	BRSTNOT14	Library was constructed using RNA isolated from breast tissue removed from a 62-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive grade 3 (of 4), nuclear grade 3 (of 3) adenocarcinoma, ductal type. Ductal carcinoma in situ, comedo type, comprised 60% of the tumor mass. Metastatic adenocarcinoma was identified in one (of 14) axillary lymph nodes with no perinodal extension. The tumor cells were strongly positive for estrogen receptors and weakly positive for progesterone receptors. Patient history included a benign colon neoplasm, hyperlipidemia, cardiac dysrhythmia, and obesity. Family history included atherosclerotic coronary artery disease, myocardial infarction, colon cancer, ovarian cancer, lung cancer, and cerebrovascular disease.
39	KIDNFET02	Library was constructed using RNA isolated from kidney tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
40	MASTTUT01	Library was constructed using RNA isolated from mast cells differentiated from treated CD34+ stem cells removed from the liver of a fetus who died at 22 weeks' gestation. The CD34+ stem cells were treated with hIL-6 and hSCF (human stem cell factor) for 18 days to induce differentiation.



Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLOCKS IMPROVED Searcher that matches a sequence against those in BLOCKS and PRINTS databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and Probability value= 1.0E-3 or less
PFAM	A Hidden Markov Models-based application useful for protein family search.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits, depending on individual protein families

Table 5 cont.

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Score= 4.0 or greater
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. <i>supra</i> ; Wisconsin Package Program Manual, version 9, page MS1-S9, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and fragments thereof.  
5
2. A substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of claim 1.
3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.  
10
4. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 3.
5. An isolated and purified polynucleotide which hybridizes under stringent  
15 conditions to the polynucleotide of claim 3.
6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 3.
- 20 7. A method for detecting a polynucleotide, the method comprising the steps of:
  - (a) hybridizing the polynucleotide of claim 6 to at least one nucleic acid in a sample, thereby forming a hybridization complex; and
  - (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide in the sample.  
25
8. The method of claim 7 further comprising amplifying the polynucleotide prior to hybridization.
9. An isolated and purified polynucleotide comprising a polynucleotide sequence  
30 selected from the group consisting of SEQ ID NO:21-40, and fragments thereof.
10. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 9.

11. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 9.
12. An expression vector comprising at least a fragment of the polynucleotide of  
5 claim 3.
13. A host cell comprising the expression vector of claim 12.
14. A method for producing a polypeptide, the method comprising the steps of:  
10 a) culturing the host cell of claim 13 under conditions suitable for the expression of the polypeptide; and  
b) recovering the polypeptide from the host cell culture.
15. A pharmaceutical composition comprising the polypeptide of claim 1 in  
15 conjunction with a suitable pharmaceutical carrier.
16. A purified antibody which specifically binds to the polypeptide of claim 1.
17. A purified agonist of the polypeptide of claim 1.  
20
18. A purified antagonist of the polypeptide of claim 1.
19. A method for treating or preventing a disorder associated with decreased expression or activity of PPRG, the method comprising administering to a subject in need of such  
25 treatment an effective amount of the pharmaceutical composition of claim 15.
20. A method for treating or preventing a disorder associated with increased expression or activity of PPRG, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 18.  
30

## SEQUENCE LISTING

&lt;110&gt; INCYTE PHARMACEUTICALS, INC.

BANDMAN, Olga  
 HILLMAN, Jennifer L.  
 BAUGHN, Mariah R.  
 AZIMZAI, Yalda  
 GUEGLER, Karl J.  
 CORLEY, Neil C.  
 YUE, Henry  
 TANG, Y. Tom  
 REDDY, Roopa  
 PATTERSON, Chandra  
 AU-YOUNG, Janice  
 SHI, Leo L.  
 LU, Dyung Aina M.

&lt;120&gt; PROTEASES AND ASSOCIATED PROTEINS

&lt;130&gt; PF-0569 PCT

&lt;140&gt; To Be Assigned

&lt;141&gt; Herewith

&lt;150&gt; 60/096,114; 60/119,768

&lt;151&gt; 1998-08-10; 1999-02-11

&lt;160&gt; 40

&lt;170&gt; PERL Program

&lt;210&gt; 1

&lt;211&gt; 206

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 1220330

&lt;400&gt; 1

Met	Pro	Ser	Arg	Arg	Arg	Asp	Ala	Ile	Lys	Val	Met	Gln	Arg	Phe
1				5					10					15
Ala	Gly	Leu	Pro	Glu	Thr	Gly	Arg	Met	Asp	Pro	Gly	Thr	Val	Ala
				20					25					30
Thr	Met	Arg	Lys	Pro	Arg	Cys	Ser	Leu	Pro	Asp	Val	Leu	Gly	Val
				35					40					45
Ala	Gly	Leu	Val	Arg	Arg	Arg	Arg	Arg	Tyr	Ala	Leu	Ser	Gly	Ser
				50					55					60
Val	Trp	Lys	Lys	Arg	Thr	Leu	Thr	Trp	Arg	Val	Arg	Ser	Phe	Pro
				65					70					75
Gln	Ser	Ser	Gln	Leu	Ser	Gln	Glu	Thr	Val	Arg	Val	Leu	Met	Ser
				80					85					90
Tyr	Ala	Leu	Met	Ala	Trp	Gly	Met	Glu	Ser	Gly	Leu	Thr	Phe	His
				95					100					105
Glu	Val	Asp	Ser	Pro	Gln	Gly	Gln	Glu	Pro	Asp	Ile	Leu	Ile	Asp

	110		115		120
Phe Ala Arg Ala	Phe His Gln Asp Ser	Tyr Pro Phe Asp Gly Leu			
	125		130		135
Gly Gly Thr Leu	Ala His Ala Phe Phe	Pro Gly Glu His Pro Ile			
	140		145		150
Ser Gly Asp Thr	His Phe Asp Asp Glu	Glu Thr Trp Thr Phe Gly			
	155		160		165
Ser Lys Ala Ser	Gln Gln Leu Glu Gln	Glu Leu Ala Gly Gly Ser			
	170		175		180
Pro Val Asp Glu	Glu Leu Gly Phe Ser	Arg Gly Trp Arg Val Asn			
	185		190		195
Pro Leu Gly Pro	Gly Ser Pro Glu Arg	Leu Ser			
	200		205		

&lt;210&gt; 2

&lt;211&gt; 754

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 1342493

&lt;400&gt; 2

Met Ala Phe Ala Ser	Trp Trp Tyr Lys Thr	His Val Ser Glu Lys
1	5	10
Thr Ser Glu Ser Pro	Ser Lys Pro Gly Glu	Lys Lys Gly Ser Asp
	20	25
Glu Lys Lys Ala Ala	Ser Leu Gly Ser Ser	Gln Ser Ser Arg Thr
	35	40
Tyr Ala Gly Gly Thr	Ala Ser Ala Thr Lys	Val Ser Ala Ser Ser
	50	55
Gly Ala Thr Ser Lys	Ser Ser Ser Met Asn	Pro Thr Glu Thr Lys
	65	70
Ala Val Lys Thr Glu	Pro Glu Lys Lys Ser	Gln Ser Thr Lys Leu
	80	85
Ser Val Val His Glu	Lys Lys Ser Gln Glu	Gly Lys Pro Lys Glu
	95	100
His Thr Glu Pro Lys	Ser Leu Pro Lys Gln	Ala Ser Asp Thr Gly
	110	115
Ser Asn Asp Ala His	Asn Lys Lys Ala Val	Ser Arg Ser Ala Glu
	125	130
Gln Gln Pro Ser Glu	Lys Ser Thr Glu Pro	Lys Thr Lys Pro Gln
	140	145
Asp Met Ile Ser Ala	Gly Gly Glu Ser Val	Ala Gly Ile Thr Ala
	155	160
Ile Ser Gly Lys Pro	Gly Asp Lys Lys Lys	Glu Lys Lys Ser Leu
	170	175
Thr Pro Ala Val Pro	Val Glu Ser Lys Pro	Asp Lys Pro Ser Gly
	185	190
Lys Ser Gly Met Asp	Ala Ala Leu Asp Asp	Leu Ile Asp Thr Leu
	200	205
Gly Gly Pro Glu Glu	Thr Glu Glu Glu Asn	Thr Thr Tyr Thr Gly
	215	220
Pro Glu Val Ser Asp	Pro Met Ser Ser Thr	Tyr Ile Glu Glu Leu
	230	235
		240

Gly	Lys	Arg	Glu	Val	Thr	Ile	Pro	Pro	Lys	Tyr	Arg	Glu	Leu	Leu
				245					250					255
Ala	Lys	Lys	Glu	Gly	Ile	Thr	Gly	Pro	Pro	Ala	Asp	Ser	Ser	Lys
				260					265					270
Pro	Ile	Gly	Pro	Asp	Asp	Ala	Ile	Asp	Ala	Leu	Ser	Ser	Asp	Phe
				275					280					285
Thr	Cys	Gly	Ser	Pro	Thr	Ala	Ala	Gly	Lys	Lys	Thr	Glu	Lys	Glu
				290					295					300
Glu	Ser	Thr	Glu	Val	Leu	Lys	Ala	Gln	Ser	Ala	Gly	Thr	Val	Arg
				305					310					315
Ser	Ala	Ala	Pro	Pro	Gln	Glu	Lys	Lys	Arg	Lys	Val	Glu	Lys	Asp
				320					325					330
Thr	Met	Ser	Asp	Gln	Ala	Leu	Glu	Ala	Leu	Ser	Ala	Ser	Leu	Gly
				335					340					345
Thr	Arg	Gln	Ala	Glu	Pro	Glu	Leu	Asp	Leu	Arg	Ser	Ile	Lys	Glu
				350					355					360
Val	Asp	Glu	Ala	Lys	Ala	Lys	Glu	Glu	Lys	Leu	Glu	Lys	Cys	Gly
				365					370					375
Glu	Asp	Asp	Glu	Thr	Ile	Pro	Ser	Glu	Tyr	Arg	Leu	Lys	Pro	Ala
				380					385					390
Thr	Asp	Lys	Asp	Gly	Lys	Pro	Leu	Leu	Pro	Glu	Pro	Glu	Glu	Lys
				395					400					405
Pro	Lys	Pro	Arg	Ser	Glu	Ser	Glu	Leu	Ile	Asp	Glu	Leu	Ser	Glu
				410					415					420
Asp	Phe	Asp	Arg	Ser	Glu	Cys	Lys	Glu	Lys	Pro	Ser	Lys	Pro	Thr
				425					430					435
Glu	Lys	Thr	Glu	Glu	Ser	Lys	Ala	Ala	Ala	Pro	Ala	Pro	Val	Ser
				440					445					450
Glu	Ala	Val	Cys	Arg	Thr	Ser	Met	Cys	Ser	Ile	Gln	Ser	Ala	Pro
				455					460					465
Pro	Glu	Pro	Ala	Thr	Leu	Lys	Gly	Thr	Val	Pro	Asp	Asp	Ala	Val
				470					475					480
Glu	Ala	Leu	Ala	Asp	Ser	Leu	Gly	Lys	Lys	Glu	Ala	Asp	Pro	Glu
				485					490					495
Asp	Gly	Lys	Pro	Val	Met	Asp	Lys	Val	Lys	Glu	Lys	Ala	Lys	Glu
				500					505					510
Glu	Asp	Arg	Glu	Lys	Leu	Gly	Glu	Lys	Glu	Glu	Thr	Ile	Pro	Pro
				515					520					525
Asp	Tyr	Arg	Leu	Glu	Glu	Val	Lys	Asp	Lys	Asp	Gly	Lys	Pro	Leu
				530					535					540
Leu	Pro	Lys	Glu	Ser	Lys	Glu	Gln	Leu	Pro	Pro	Met	Ser	Glu	Asp
				545					550					555
Phe	Leu	Leu	Asp	Ala	Leu	Ser	Glu	Asp	Phe	Ser	Gly	Pro	Gln	Asn
				560					565					570
Ala	Ser	Ser	Leu	Lys	Phe	Glu	Asp	Ala	Lys	Leu	Ala	Ala	Ala	Ile
				575					580					585
Ser	Glu	Val	Val	Ser	Gln	Thr	Pro	Ala	Ser	Thr	Thr	Gln	Ala	Gly
				590					595					600
Ala	Pro	Pro	Arg	Asp	Thr	Ser	Gln	Ser	Asp	Lys	Asp	Leu	Asp	Asp
				605					610					615
Ala	Leu	Asp	Lys	Leu	Ser	Asp	Ser	Leu	Gly	Gln	Arg	Gln	Pro	Asp
				620					625					630
Pro	Asp	Glu	Asn	Lys	Pro	Met	Glu	Asp	Lys	Val	Lys	Glu	Lys	Ala
				635					640					645
Lys	Ala	Glu	His	Arg	Asp	Lys	Leu	Gly	Glu	Arg	Asp	Asp	Thr	Ile
				650					655					660
Pro	Pro	Glu	Tyr	Arg	His	Leu	Leu	Asp	Asp	Asn	Gly	Gln	Asp	Lys

	665		670		675
Pro Val Lys Pro	Pro Thr Lys Lys Ser	Glu Asp Ser Lys Lys	Pro		
	680		685		690
Ala Asp Asp Gln	Asp Pro Ile Asp Ala	Leu Ser Gly Asp Leu	Asp		
	695		700		705
Ser Cys Pro Ser	Thr Thr Glu Thr Ser	Gln Asn Thr Ala Lys	Asp		
	710		715		720
Lys Cys Lys Lys	Ala Ala Ser Ser Ser	Lys Ala Pro Lys Asn	Gly		
	725		730		735
Gly Lys Ala Lys	Asp Ser Ala Lys Thr	Thr Glu Glu Thr Ser	Lys		
	740		745		750
Pro Lys Asp Asp					

&lt;210&gt; 3

&lt;211&gt; 308

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 1698270

&lt;400&gt; 3

Met Gly Glu Ile Lys Val Ser Pro Asp Tyr Asn Trp Phe Arg Gly		
1	5	10
Thr Val Pro Leu Lys Lys Ile Ile Val Asp Asp Asp Asp Ser Lys		
	20	25
Ile Trp Ser Leu Tyr Asp Ala Gly Pro Arg Ser Ile Arg Cys Pro		
	35	40
Leu Ile Phe Leu Pro Pro Val Ser Gly Thr Ala Asp Val Phe Phe		
	50	55
Arg Gln Ile Leu Ala Leu Thr Gly Trp Gly Tyr Arg Val Ile Ala		
	65	70
Leu Gln Tyr Pro Val Tyr Trp Asp His Leu Glu Phe Cys Asp Gly		
	80	85
Phe Arg Lys Leu Leu Asp His Leu Gln Leu Asp Lys Val His Leu		
	95	100
Phe Gly Ala Ser Leu Gly Gly Phe Leu Ala Gln Lys Phe Ala Glu		
	110	115
Tyr Thr His Lys Ser Pro Arg Val His Ser Leu Ile Leu Cys Asn		
	125	130
Ser Phe Ser Asp Thr Ser Ile Phe Asn Gln Thr Trp Thr Ala Asn		
	140	145
Ser Phe Trp Leu Met Pro Ala Phe Met Leu Lys Lys Ile Val Leu		
	155	160
Gly Asn Phe Ser Ser Gly Pro Val Asp Pro Met Met Ala Asp Ala		
	170	175
Ile Asp Phe Met Val Asp Arg Leu Glu Ser Leu Gly Gln Ser Glu		
	185	190
Leu Ala Ser Arg Leu Thr Leu Asn Cys Gln Asn Ser Tyr Val Glu		
	200	205
Pro His Lys Ile Arg Asp Ile Pro Val Thr Ile Met Asp Val Phe		
	215	220
Asp Gln Ser Ala Leu Ser Thr Glu Ala Lys Glu Glu Met Tyr Lys		
	230	235
Leu Tyr Pro Asn Ala Arg Arg Ala His Leu Lys Thr Gly Gly Asn		



	245		250		255
Phe Pro Tyr Leu Cys Arg Ser Ala Glu Val Asn Leu Tyr Val Gln					
	260		265		270
Ile His Leu Leu Gln Phe His Gly Thr Lys Tyr Ala Ala Ile Asp					
	275		280		285
Pro Ser Met Val Ser Ala Glu Glu Leu Glu Val Gln Lys Gly Ser					
	290		295		300
Leu Gly Ile Ser Gln Glu Glu Gln					
	305				

<210> 4  
 <211> 164  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte Clone No: 2012492

<400> 4	
Met Lys Thr Gln Asp Gly Gly Ile His Ser Glu Gly Ala Ala Ala	
1 5 10 15	
Glu His Ser Lys Phe Gly Asn His Gln Lys Gly Trp Pro Leu Phe	
20 25 30	
Asn Met Gly Ser Ser Gly Leu Leu Ser Leu Leu Val Leu Phe Val	
35 40 45	
Leu Leu Ala Asn Val Gln Gly Pro Gly Leu Thr Asp Trp Leu Phe	
50 55 60	
Pro Arg Arg Cys Pro Lys Ile Arg Glu Glu Cys Glu Phe Gln Glu	
65 70 75	
Arg Asp Val Cys Thr Lys Asp Arg Gln Cys Gln Asp Asn Lys Lys	
80 85 90	
Cys Cys Val Phe Ser Cys Gly Lys Lys Cys Leu Asp Leu Lys Gln	
95 100 105	
Asp Val Cys Glu Met Pro Lys Glu Thr Gly Pro Cys Leu Ala Tyr	
110 115 120	
Phe Leu His Trp Trp Tyr Asp Lys Lys Asp Asn Thr Cys Ser Met	
125 130 135	
Phe Val Tyr Gly Gly Cys Gln Gly Asn Asn Asn Asn Phe Gln Ser	
140 145 150	
Lys Ala Asn Cys Leu Asn Thr Cys Lys Asn Lys Arg Phe Pro	
155 160	

<210> 5  
 <211> 565  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte Clone No: 2309875

<400> 5	
Met Pro Gln Ala Ser Glu His Arg Leu Gly Arg Thr Arg Glu Pro	
1 5 10 15	
Pro Val Asn Ile Gln Pro Arg Val Gly Ser Lys Leu Pro Phe Ala	

20	25	30
Pro Arg Ala Arg Ser Lys Glu Arg Arg Asn Pro Ala Ser Gly Pro		
35	40	45
Asn Pro Met Leu Arg Pro Leu Pro Pro Arg Pro Gly Leu Pro Asp		
50	55	60
Glu Arg Leu Lys Lys Leu Glu Leu Gly Arg Gly Arg Thr Ser Gly		
65	70	75
Pro Arg Pro Arg Gly Pro Leu Arg Ala Asp His Gly Val Pro Leu		
80	85	90
Pro Gly Ser Pro Pro Pro Thr Val Ala Leu Pro Leu Pro Ser Arg		
95	100	105
Thr Asn Leu Ala Arg Ser Lys Ser Val Ser Ser Gly Asp Leu Arg		
110	115	120
Pro Met Gly Ile Ala Leu Gly Gly His Arg Gly Thr Gly Glu Leu		
125	130	135
Gly Ala Ala Leu Ser Arg Leu Ala Leu Arg Pro Glu Pro Pro Thr		
140	145	150
Leu Arg Arg Ser Thr Ser Leu Arg Arg Leu Gly Gly Phe Pro Gly		
155	160	165
Pro Pro Thr Leu Phe Ser Ile Arg Thr Glu Pro Pro Ala Ser His		
170	175	180
Gly Ser Phe His Met Ile Ser Ala Arg Ser Ser Glu Pro Phe Tyr		
185	190	195
Ser Asp Asp Lys Met Ala His His Thr Leu Leu Leu Gly Ser Gly		
200	205	210
His Val Gly Leu Arg Asn Leu Gly Asn Thr Cys Phe Leu Asn Ala		
215	220	225
Val Leu Gln Cys Leu Ser Ser Thr Arg Pro Leu Arg Asp Phe Cys		
230	235	240
Leu Arg Arg Asp Phe Arg Gln Glu Val Pro Gly Gly Gly Arg Ala		
245	250	255
Gln Glu Leu Thr Glu Ala Phe Ala Asp Val Ile Gly Ala Leu Trp		
260	265	270
His Pro Asp Ser Cys Glu Ala Val Asn Pro Thr Arg Phe Arg Ala		
275	280	285
Val Phe Gln Lys Tyr Val Pro Ser Phe Ser Gly Tyr Ser Gln Gln		
290	295	300
Asp Ala Gln Glu Phe Leu Lys Leu Leu Met Glu Arg Leu His Leu		
305	310	315
Glu Ile Asn Arg Arg Gly Arg Arg Ala Pro Pro Ile Leu Ala Asn		
320	325	330
Gly Pro Val Pro Ser Pro Pro Arg Arg Gly Gly Ala Leu Leu Glu		
335	340	345
Glu Pro Glu Leu Ser Asp Asp Asp Arg Ala Asn Leu Met Trp Lys		
350	355	360
Arg Tyr Leu Glu Arg Glu Asp Ser Lys Ile Val Asp Leu Phe Val		
365	370	375
Gly Gln Leu Lys Ser Cys Leu Lys Cys Gln Ala Cys Gly Tyr Arg		
380	385	390
Ser Thr Thr Phe Glu Val Phe Cys Asp Leu Ser Leu Pro Ile Pro		
395	400	405
Lys Lys Gly Phe Ala Gly Gly Lys Val Ser Leu Arg Asp Cys Phe		
410	415	420
Asn Leu Phe Thr Lys Glu Glu Glu Leu Glu Ser Glu Asn Ala Pro		
425	430	435
Val Cys Asp Arg Cys Arg Gln Lys Thr Arg Ser Thr Lys Lys Leu		
440	445	450

Thr	Val	Gln	Arg	Phe	Pro	Arg	Ile	Leu	Val	Leu	His	Leu	Asn	Arg
				455					460					465
Phe	Ser	Ala	Ser	Arg	Gly	Ser	Ile	Lys	Lys	Ser	Ser	Val	Gly	Val
				470					475					480
Asp	Phe	Pro	Leu	Gln	Arg	Leu	Ser	Leu	Gly	Asp	Phe	Ala	Ser	Asp
				485					490					495
Lys	Ala	Gly	Ser	Pro	Val	Tyr	Gln	Leu	Tyr	Ala	Leu	Cys	Asn	His
				500					505					510
Ser	Gly	Ser	Val	His	Tyr	Gly	His	Tyr	Thr	Ala	Leu	Cys	Arg	Cys
				515					520					525
Gln	Thr	Gly	Trp	His	Val	Tyr	Asn	Asp	Ser	Arg	Val	Ser	Pro	Val
				530					535					540
Ser	Glu	Asn	Gln	Val	Ala	Ser	Ser	Glu	Gly	Tyr	Val	Leu	Phe	Tyr
				545					550					555
Gln	Leu	Met	Gln	Glu	Pro	Pro	Arg	Cys	Leu					
				560					565					

&lt;210&gt; 6

&lt;211&gt; 421

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2479394

&lt;400&gt; 6

Met	Arg	Trp	Ile	Leu	Phe	Ile	Gly	Ala	Leu	Ile	Gly	Ser	Ser	Ile
1				5					10					15
Cys	Gly	Gln	Glu	Lys	Phe	Phe	Gly	Asp	Gln	Val	Leu	Arg	Ile	Asn
				20					25					30
Val	Arg	Asn	Gly	Asp	Glu	Ile	Ser	Lys	Leu	Ser	Gln	Leu	Val	Asn
				35					40					45
Ser	Asn	Asn	Leu	Lys	Leu	Asn	Phe	Trp	Lys	Ser	Pro	Ser	Ser	Phe
				50					55					60
Asn	Arg	Pro	Val	Asp	Val	Leu	Val	Pro	Ser	Val	Ser	Leu	Gln	Ala
				65					70					75
Phe	Lys	Ser	Phe	Leu	Arg	Ser	Gln	Gly	Leu	Glu	Tyr	Ala	Val	Thr
				80					85					90
Ile	Glu	Asp	Leu	Gln	Ala	Leu	Leu	Asp	Asn	Glu	Asp	Asp	Glu	Met
				95					100					105
Gln	His	Asn	Glu	Gly	Gln	Glu	Arg	Ser	Ser	Asn	Asn	Phe	Asn	Tyr
				110					115					120
Gly	Ala	Tyr	His	Ser	Leu	Glu	Ala	Ile	Tyr	His	Glu	Met	Asp	Asn
				125					130					135
Ile	Ala	Ala	Asp	Phe	Pro	Asp	Leu	Ala	Arg	Arg	Val	Lys	Ile	Gly
				140					145					150
His	Ser	Phe	Glu	Asn	Arg	Pro	Met	Tyr	Val	Leu	Lys	Phe	Ser	Thr
				155					160					165
Gly	Lys	Gly	Val	Arg	Arg	Pro	Ala	Val	Trp	Leu	Asn	Ala	Gly	Ile
				170					175					180
His	Ser	Arg	Glu	Trp	Ile	Ser	Gln	Ala	Thr	Ala	Ile	Trp	Thr	Ala
				185					190					195
Arg	Lys	Ile	Val	Ser	Asp	Tyr	Gln	Arg	Asp	Pro	Ala	Ile	Thr	Ser
				200					205					210
Ile	Leu	Glu	Lys	Met	Asp	Ile	Phe	Leu	Leu	Pro	Val	Ala	Asn	Pro

Asp Gly Tyr Val	215	Tyr Thr Gln Thr Gln	220	Asn Arg Leu Trp Arg Lys	225
Thr Arg Ser Arg	230	Asn Pro Gly Ser Ser	235	Cys Ile Gly Ala Asp Pro	240
Asn Arg Asn Trp	245	Asn Ala Ser Phe Ala	250	Gly Lys Gly Ala Ser Asp	255
Asn Pro Cys Ser	260	Glu Val Tyr His Gly	265	Pro His Ala Asn Ser Glu	270
Val Glu Val Lys	275	Ser Val Val Asp Phe	280	Ile Gln Lys His Gly Asn	285
Phe Lys Gly Phe	290	Ile Asp Leu His Ser	295	Tyr Ser Gln Leu Leu Met	300
Tyr Pro Tyr Gly	305	Tyr Ser Val Lys Lys	310	Ala Pro Asp Ala Glu Glu	315
Leu Asp Lys Val	320	Ala Arg Leu Ala Ala	325	Lys Ala Leu Ala Ser Val	330
Ser Gly Thr Glu	335	Tyr Gln Val Gly Pro	340	Thr Cys Thr Thr Val Tyr	345
Pro Ala Ser Gly	350	Ser Ser Ile Asp Trp	355	Ala Tyr Asp Asn Gly Ile	360
Lys Phe Ala Phe	365	Thr Phe Glu Leu Arg	370	Asp Thr Gly Thr Tyr Gly	375
Phe Leu Leu Pro	380	Ala Asn Gln Ile Ile	385	Pro Thr Ala Glu Glu Thr	390
Trp Leu Gly Leu	395	Lys Thr Ile Met Glu	400	His Val Arg Asp Asn Leu	405
Tyr	410		415		420

&lt;210&gt; 7

&lt;211&gt; 666

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2613215

&lt;400&gt; 7

Met Ala Ala Ser Arg Lys Pro Pro Arg Val Arg Val Asn His Gln	1	5	10	15
Asp Phe Gln Leu Arg Asn Leu Arg Ile Ile Glu Pro Asn Glu Val	20	25	30	
Thr His Ser Gly Asp Thr Gly Val Glu Thr Asp Gly Arg Met Pro	35	40	45	
Pro Lys Val Thr Ser Glu Leu Leu Arg Gln Leu Arg Gln Ala Met	50	55	60	
Arg Asn Ser Glu Tyr Val Thr Glu Pro Ile Gln Ala Tyr Ile Ile	65	70	75	
Pro Ser Gly Asp Ala His Gln Ser Glu Tyr Ile Ala Pro Cys Asp	80	85	90	
Cys Arg Arg Ala Phe Val Ser Gly Phe Asp Gly Ser Ala Gly Thr	95	100	105	
Ala Ile Ile Thr Glu Glu His Ala Ala Met Trp Thr Asp Gly Arg	110	115	120	
Tyr Phe Leu Gln Ala Ala Lys Gln Met Asp Ser Asn Trp Thr Leu				

	125		130		135
Met Lys Met Gly	Leu Lys Asp Thr Pro	Gln Glu Asp Trp	Leu		
	140		145		150
Val Ser Val Leu	Pro Glu Gly Ser Arg	Val Gly Val Asp Pro	Leu		
	155		160		165
Ile Ile Pro Thr	Asp Tyr Trp Lys Lys	Met Ala Lys Val Leu	Arg		
	170		175		180
Ser Ala Gly His	His Leu Ile Pro Val	Lys Glu Asn Leu Val	Asp		
	185		190		195
Lys Ile Trp Thr	Asp Arg Pro Glu Arg	Pro Cys Lys Pro Leu	Leu		
	200		205		210
Thr Leu Gly Leu	Asp Tyr Thr Gly Ile	Ser Trp Lys Asp Lys	Val		
	215		220		225
Ala Asp Leu Arg	Leu Lys Met Ala Glu	Arg Asn Val Met Trp	Phe		
	230		235		240
Val Val Thr Ala	Leu Asp Glu Ile Ala	Trp Leu Phe Asn Leu	Arg		
	245		250		255
Gly Ser Asp Val	Glu His Asn Pro Val	Phe Phe Ser Tyr Ala	Ile		
	260		265		270
Ile Gly Leu Glu	Thr Ile Met Leu Phe	Ile Asp Gly Asp Arg	Ile		
	275		280		285
Asp Ala Pro Ser	Val Lys Glu His Leu	Leu Leu Asp Leu Gly	Leu		
	290		295		300
Glu Ala Glu Tyr	Arg Ile Gln Val His	Pro Tyr Lys Ser Ile	Leu		
	305		310		315
Ser Glu Leu Lys	Ala Leu Cys Ala Asp	Leu Ser Pro Arg Glu	Lys		
	320		325		330
Val Trp Val Ser	Asp Lys Ala Ser Tyr	Ala Val Ser Glu Thr	Ile		
	335		340		345
Pro Lys Asp His	Arg Cys Cys Met Pro	Tyr Thr Pro Ile Cys	Ile		
	350		355		360
Ala Lys Ala Val	Lys Asn Ser Ala Glu	Ser Glu Gly Met Arg	Arg		
	365		370		375
Ala His Ile Lys	Asp Ala Val Ala Leu	Cys Glu Leu Phe Asn	Trp		
	380		385		390
Leu Glu Lys Glu	Val Pro Lys Gly Gly	Val Thr Glu Ile Ser	Ala		
	395		400		405
Ala Asp Lys Ala	Glu Glu Phe Arg Arg	Gln Gln Ala Asp Phe	Val		
	410		415		420
Asp Leu Ser Phe	Pro Thr Ile Ser Ser	Thr Gly Pro Asn Gly	Ala		
	425		430		435
Ile Ile His Tyr	Ala Pro Val Pro Glu	Thr Asn Arg Thr Leu	Ser		
	440		445		450
Leu Asp Glu Val	Tyr Leu Ile Asp Ser	Gly Ala Gln Tyr Lys	Asp		
	455		460		465
Gly Thr Thr Asp	Val Thr Arg Thr Met	His Phe Gly Thr Pro	Thr		
	470		475		480
Ala Tyr Glu Lys	Glu Cys Phe Thr Tyr	Val Leu Lys Gly His	Ile		
	485		490		495
Ala Val Ser Ala	Ala Val Phe Pro Thr	Gly Thr Lys Gly His	Leu		
	500		505		510
Leu Asp Ser Phe	Ala Arg Ser Ala Leu	Trp Asp Ser Gly Leu	Asp		
	515		520		525
Tyr Leu His Gly	Thr Gly His Gly Val	Gly Ser Phe Leu Asn	Val		
	530		535		540
His Glu Gly Pro	Cys Gly Ile Ser Tyr	Lys Thr Phe Ser Asp	Glu		
	545		550		555

```

Pro Leu Glu Ala Gly Met Ile Val Thr Asp Glu Pro Gly Tyr Tyr
                    560                    565                    570
Glu Asp Gly Ala Phe Gly Ile Arg Ile Glu Asn Val Val Leu Val
                    575                    580                    585
Val Pro Val Lys Thr Lys Tyr Asn Phe Asn Asn Arg Gly Ser Leu
                    590                    595                    600
Thr Phe Glu Pro Leu Thr Leu Val Pro Ile Gln Thr Lys Met Ile
                    605                    610                    615
Asp Val Asp Ser Leu Thr Asp Lys Glu Cys Asp Trp Leu Asn Asn
                    620                    625                    630
Tyr His Leu Thr Cys Arg Asp Val Ile Gly Lys Glu Leu Gln Lys
                    635                    640                    645
Gln Gly Arg Gln Glu Ala Leu Glu Trp Leu Ile Arg Glu Thr Gln
                    650                    655                    660
Pro Ile Ser Lys Gln His
                    665

```

&lt;210&gt; 8

&lt;211&gt; 952

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 001528

&lt;400&gt; 8

```

Met Ala Glu Gly Gly Ala Ala Asp Leu Asp Thr Gln Arg Ser Asp
  1          5          10          15
Ile Ala Thr Leu Leu Lys Thr Ser Leu Arg Lys Gly Asp Thr Trp
          20          25          30
Tyr Leu Val Asp Ser Arg Trp Phe Lys Gln Trp Lys Lys Tyr Val
          35          40          45
Gly Phe Asp Ser Trp Asp Lys Tyr Gln Met Gly Asp Gln Asn Val
          50          55          60
Tyr Pro Gly Pro Ile Asp Asn Ser Gly Leu Leu Lys Asp Gly Asp
          65          70          75
Ala Gln Ser Leu Lys Glu His Leu Ile Asp Glu Leu Asp Tyr Ile
          80          85          90
Leu Leu Pro Thr Glu Gly Trp Asn Lys Leu Val Ser Trp Tyr Thr
          95          100          105
Leu Met Glu Gly Gln Glu Pro Ile Ala Arg Lys Val Val Glu Gln
          110          115          120
Gly Met Phe Val Lys Arg Cys Lys Val Glu Val Tyr Leu Thr Glu
          125          130          135
Leu Lys Leu Cys Glu Asn Gly Asn Met Asn Asn Val Val Thr Arg
          140          145          150
Arg Phe Ser Lys Ala Asp Thr Ile Asp Thr Ile Glu Lys Glu Ile
          155          160          165
Arg Lys Ile Phe Ser Ile Pro Asp Glu Lys Glu Thr Arg Leu Trp
          170          175          180
Asn Lys Tyr Met Ser Asn Thr Phe Glu Pro Leu Asn Lys Pro Asp
          185          190          195
Ser Thr Ile Gln Asp Ala Gly Leu Tyr Gln Gly Gln Val Leu Val
          200          205          210
Ile Glu Gln Lys Asn Glu Asp Gly Thr Arg Pro Arg Gly Pro Ser

```

Thr Pro Asn Val	215	220	225
Lys Asn Ser Asn Tyr		Cys Leu Pro Ser Tyr	Thr
230		235	240
Ala Tyr Lys Asn Tyr Asp Tyr Ser Glu		Pro Gly Arg Asn Asn	Glu
245		250	255
Gln Pro Gly Leu Cys Gly Leu Ser Asn		Leu Gly Asn Thr Cys	Phe
260		265	270
Met Asn Ser Ala Ile Gln Cys Leu Ser		Asn Thr Pro Pro Leu	Thr
275		280	285
Glu Tyr Phe Leu Asn Asp Lys Tyr Gln		Glu Glu Leu Asn Phe	Asp
290		295	300
Asn Pro Leu Gly Met Arg Gly Glu Ile		Ala Lys Ser Tyr Ala	Glu
305		310	315
Leu Ile Lys Gln Met Trp Ser Gly Lys		Phe Ser Tyr Val Thr	Pro
320		325	330
Arg Ala Phe Lys Thr Gln Val Gly Arg		Phe Ala Pro Gln Phe	Ser
335		340	345
Gly Tyr Gln Gln Gln Asp Cys Gln Glu		Leu Leu Ala Phe Leu	Leu
350		355	360
Asp Gly Leu His Glu Asp Leu Asn Arg		Ile Arg Lys Lys Pro	Tyr
365		370	375
Ile Gln Leu Lys Asp Ala Asp Gly Arg		Pro Asp Lys Val Val	Ala
380		385	390
Glu Glu Ala Trp Glu Asn His Leu Lys		Arg Asn Asp Ser Ile	Ile
395		400	405
Val Asp Ile Phe His Gly Leu Phe Lys		Ser Thr Leu Val Cys	Pro
410		415	420
Glu Cys Ala Lys Ile Ser Val Thr Phe		Asp Pro Phe Cys Tyr	Leu
425		430	435
Thr Leu Pro Leu Pro Met Lys Lys Glu		Arg Thr Leu Glu Val	Tyr
440		445	450
Leu Val Arg Met Asp Pro Leu Thr Lys		Pro Met Gln Tyr Lys	Val
455		460	465
Val Val Pro Lys Ile Gly Asn Ile Leu		Asp Leu Cys Thr Ala	Leu
470		475	480
Ser Ala Leu Ser Gly Ile Pro Ala Asp		Lys Met Ile Val Thr	Asp
485		490	495
Ile Tyr Asn His Arg Phe His Arg Ile		Phe Ala Met Asp Glu	Asn
500		505	510
Leu Ser Ser Ile Met Glu Arg Asp Asp		Ile Tyr Val Phe Glu	Ile
515		520	525
Asn Ile Asn Arg Thr Glu Asp Thr Glu		His Val Ile Ile Pro	Val
530		535	540
Cys Leu Arg Glu Lys Phe Arg His Ser		Ser Tyr Thr His His	Thr
545		550	555
Gly Ser Ser Leu Phe Gly Gln Pro Phe		Leu Met Ala Val Pro	Arg
560		565	570
Asn Asn Thr Glu Asp Lys Leu Tyr Asn		Leu Leu Leu Leu Arg	Met
575		580	585
Cys Arg Tyr Val Lys Ile Ser Thr Glu		Thr Glu Glu Thr Glu	Gly
590		595	600
Ser Leu His Cys Cys Lys Asp Gln Asn		Ile Asn Gly Asn Gly	Pro
605		610	615
Asn Gly Ile His Glu Glu Gly Ser Pro		Ser Glu Met Glu Thr	Asp
620		625	630
Glu Pro Asp Asp Glu Ser Ser Gln Asp		Gln Glu Leu Pro Ser	Glu
635		640	645

Asn Glu Asn Ser Gln Ser Glu Asp Ser Val Gly Gly Asp Asn Asp  
 650 655 660  
 Ser Glu Asn Gly Leu Cys Thr Glu Asp Thr Cys Lys Gly Gln Leu  
 665 670 675  
 Thr Gly His Lys Lys Arg Leu Phe Thr Phe Gln Phe Asn Asn Leu  
 680 685 690  
 Gly Asn Thr Asp Ile Asn Tyr Ile Lys Asp Asp Thr Arg His Ile  
 695 700 705  
 Arg Phe Asp Asp Arg Gln Leu Arg Leu Asp Glu Arg Ser Phe Leu  
 710 715 720  
 Ala Leu Asp Trp Asp Pro Asp Leu Lys Lys Arg Tyr Phe Asp Glu  
 725 730 735  
 Asn Ala Ala Glu Asp Phe Glu Lys His Glu Ser Val Glu Tyr Lys  
 740 745 750  
 Pro Pro Lys Lys Pro Phe Val Lys Leu Lys Asp Cys Ile Glu Leu  
 755 760 765  
 Phe Thr Thr Lys Glu Lys Leu Gly Ala Glu Asp Pro Trp Tyr Cys  
 770 775 780  
 Pro Asn Cys Lys Glu His Gln Gln Ala Thr Lys Lys Leu Asp Leu  
 785 790 795  
 Trp Ser Leu Pro Pro Val Leu Val Val His Leu Lys Arg Phe Ser  
 800 805 810  
 Tyr Ser Arg Tyr Met Arg Asp Lys Leu Asp Thr Leu Val Asp Phe  
 815 820 825  
 Pro Ile Asn Asp Leu Asp Met Ser Glu Phe Leu Ile Asn Pro Asn  
 830 835 840  
 Ala Gly Pro Cys Arg Tyr Asn Leu Ile Ala Val Ser Asn His Tyr  
 845 850 855  
 Gly Gly Met Gly Gly Gly His Tyr Thr Ala Phe Ala Lys Asn Lys  
 860 865 870  
 Asp Asp Gly Lys Trp Tyr Tyr Phe Asp Asp Ser Ser Val Ser Thr  
 875 880 885  
 Ala Ser Glu Asp Gln Ile Val Ser Lys Ala Ala Tyr Val Leu Phe  
 890 895 900  
 Tyr Gln Arg Gln Asp Thr Phe Ser Gly Thr Gly Phe Phe Pro Leu  
 905 910 915  
 Asp Arg Glu Thr Lys Gly Ala Ser Ala Ala Thr Gly Ile Pro Leu  
 920 925 930  
 Glu Ser Asp Glu Asp Ser Asn Asp Asn Asp Asn Asp Ile Glu Asn  
 935 940 945  
 Glu Asn Cys Met His Thr Asn  
 950

&lt;210&gt; 9

&lt;211&gt; 166

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 998626

&lt;400&gt; 9

Met Leu His Pro Glu Thr Ser Pro Gly Arg Gly His Leu Leu Ala  
 1 5 10 15  
 Val Leu Leu Ala Leu Leu Gly Thr Ala Trp Ala Glu Val Trp Pro



<400> 10														
Met	Arg	Lys	Pro	Ala	Ala	Gly	Phe	Leu	Pro	Ser	Leu	Leu	Lys	Val
1				5					10					15
Leu	Leu	Leu	Pro	Leu	Ala	Pro	Ala	Ala	Ala	Gln	Asp	Ser	Thr	Gln
				20					25					30
Ala	Ser	Thr	Pro	Gly	Ser	Pro	Leu	Ser	Pro	Thr	Glu	Tyr	Glu	Arg
				35					40					45
Phe	Phe	Ala	Leu	Leu	Thr	Pro	Thr	Trp	Lys	Ala	Glu	Thr	Thr	Cys
				50					55					60
Arg	Leu	Arg	Ala	Thr	His	Gly	Cys	Arg	Asn	Pro	Thr	Leu	Val	Gln
				65					70					75
Leu	Asp	Gln	Tyr	Glu	Asn	His	Gly	Leu	Val	Pro	Asp	Gly	Ala	Val
				80					85					90
Cys	Ser	Asn	Leu	Pro	Tyr	Ala	Ser	Trp	Phe	Glu	Ser	Phe	Cys	Gln
				95					100					105
Phe	Thr	His	Tyr	Arg	Cys	Ser	Asn	His	Val	Tyr	Tyr	Ala	Lys	Arg
				110					115					120
Val	Leu	Cys	Ser	Gln	Pro	Val	Ser	Ile	Leu	Ser	Pro	Asn	Thr	Leu
				125					130					135
Lys	Glu	Ile	Glu	Ala	Ser	Ala	Glu	Val	Ser	Pro	Thr	Thr	Met	Thr
				140					145					150
Ser	Pro	Ile	Ser	Pro	His	Phe	Thr	Val	Thr	Glu	Arg	Gln	Thr	Phe
				155					160					165
Gln	Pro	Trp	Pro	Glu	Arg	Leu	Ser	Asn	Asn	Val	Glu	Glu	Leu	Leu
				170					175					180
Gln	Ser	Ser	Leu	Ser	Leu	Gly	Gly	Gln	Glu	Gln	Ala	Pro	Glu	His

	185		190		195
Lys Gln Glu Gln Gly Val Glu His Arg		Gln Glu Pro Thr Gln Glu			
	200		205		210
His Lys Gln Glu Glu Gly Gln Lys Gln		Glu Glu Gln Glu Glu Glu			
	215		220		225
Gln Glu Glu Glu Gly Lys Gln Glu Glu		Gly Gln Gly Thr Lys Glu			
	230		235		240
Gly Arg Glu Ala Val Ser Gln Leu Gln		Thr Asp Ser Glu Pro Lys			
	245		250		255
Phe His Ser Glu Ser Leu Ser Ser Asn		Pro Ser Ser Phe Ala Pro			
	260		265		270
Arg Val Arg Glu Val Glu Ser Thr Pro		Met Ile Met Glu Asn Ile			
	275		280		285
Gln Glu Leu Ile Arg Ser Ala Gln Glu		Ile Asp Glu Met Asn Glu			
	290		295		300
Ile Tyr Asp Glu Asn Ser Tyr Trp Arg		Asn Gln Asn Pro Gly Ser			
	305		310		315
Leu Leu Gln Leu Pro His Thr Glu Ala		Leu Leu Val Leu Cys Tyr			
	320		325		330
Ser Ile Val Glu Asn Thr Cys Ile Ile		Thr Pro Thr Ala Lys Ala			
	335		340		345
Trp Lys Tyr Met Glu Glu Glu Ile Leu		Gly Phe Gly Lys Ser Val			
	350		355		360
Cys Asp Ser Leu Gly Arg Arg His Met		Ser Thr Cys Ala Leu Cys			
	365		370		375
Asp Phe Cys Ser Leu Lys Leu Glu Gln		Cys His Ser Glu Ala Ser			
	380		385		390
Leu Gln Arg Gln Gln Cys Asp Thr Ser		His Lys Thr Pro Phe Val			
	395		400		405
Ser Pro Leu Leu Ala Ser Gln Ser Leu		Ser Ile Gly Asn Gln Val			
	410		415		420
Gly Ser Pro Glu Ser Gly Arg Phe Tyr		Gly Leu Asp Leu Tyr Gly			
	425		430		435
Gly Leu His Met Asp Phe Trp Cys Ala		Arg Leu Ala Thr Lys Gly			
	440		445		450
Cys Glu Asp Val Arg Val Ser Gly Trp		Leu Gln Thr Glu Phe Leu			
	455		460		465
Ser Phe Gln Asp Gly Asp Phe Pro Thr		Lys Ile Cys Asp Thr Asp			
	470		475		480
Tyr Ile Gln Tyr Pro Asn Tyr Cys Ser		Phe Lys Ser Gln Gln Cys			
	485		490		495
Leu Met Arg Asn Arg Asn Arg Lys Val		Ser Arg Met Arg Cys Leu			
	500		505		510
Gln Asn Glu Thr Tyr Ser Ala Leu Ser		Pro Gly Lys Ser Glu Asp			
	515		520		525
Val Val Leu Arg Trp Ser Gln Glu Phe		Ser Thr Leu Thr Leu Gly			
	530		535		540
Gln Phe Gly					

&lt;210&gt; 11

&lt;211&gt; 83

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

<223> Incyte Clone No: 1444055

<400> 11

Met	Ile	Gly	Trp	Asp	Ser	Leu	Arg	Leu	Ile	Leu	Gly	Asn	Thr	Asp
1				5					10					15
Asn	Val	Ser	Arg	Arg	Asp	Ser	Thr	Arg	Gly	Ser	Ile	Phe	Ile	Thr
				20					25					30
Gln	Leu	Ile	Ala	Cys	Phe	Gln	Arg	Tyr	Ser	Trp	Arg	Cys	His	Leu
				35					40					45
Glu	Glu	Val	Phe	Trp	Lys	Val	Gln	Gln	Ala	Phe	Glu	Ser	Pro	Glu
				50					55					60
Ala	Thr	Val	Gln	Met	Pro	Thr	Ile	Glu	Arg	Val	Ser	Met	Thr	Arg
				65					70					75
Tyr	Phe	Tyr	Leu	Phe	Pro	Gly	Asn							
				80										

<210> 12

<211> 648

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte Clone No: 1650177

<400> 12

Met	Leu	Gly	Ser	Leu	Val	Leu	Arg	Arg	Lys	Ala	Leu	Ala	Pro	Arg
1				5					10					15
Leu	Leu	Leu	Arg	Leu	Leu	Arg	Ser	Pro	Thr	Leu	Arg	Gly	His	Gly
				20					25					30
Gly	Ala	Ser	Gly	Arg	Asn	Val	Thr	Thr	Gly	Ser	Leu	Gly	Glu	Pro
				35					40					45
Gln	Trp	Leu	Arg	Val	Ala	Thr	Gly	Gly	Arg	Pro	Gly	Thr	Ser	Pro
				50					55					60
Ala	Leu	Phe	Ser	Gly	Arg	Gly	Ala	Ala	Thr	Gly	Gly	Arg	Gln	Gly
				65					70					75
Gly	Arg	Phe	Asp	Thr	Lys	Cys	Leu	Ala	Ala	Ala	Thr	Trp	Gly	Arg
				80					85					90
Leu	Pro	Gly	Pro	Glu	Glu	Thr	Leu	Pro	Gly	Gln	Asp	Ser	Trp	Asn
				95					100					105
Gly	Val	Pro	Ser	Arg	Ala	Gly	Leu	Gly	Met	Cys	Ala	Leu	Ala	Ala
				110					115					120
Ala	Leu	Val	Val	His	Cys	Tyr	Ser	Lys	Ser	Pro	Ser	Asn	Lys	Asp
				125					130					135
Ala	Ala	Leu	Leu	Glu	Ala	Ala	Arg	Ala	Asn	Asn	Met	Gln	Glu	Val
				140					145					150
Ser	Ser	Val	Val	Gln	Val	Leu	Leu	Ala	Ala	Gly	Ala	Asp	Pro	Asn
				155					160					165
Leu	Gly	Asp	Asp	Phe	Ser	Ser	Val	Phe	Lys	Thr	Ala	Lys	Glu	Gln
				170					175					180
Gly	Ile	His	Ser	Leu	Glu	Val	Leu	Ile	Thr	Arg	Glu	Asp	Asp	Phe
				185					190					195
Asn	Asn	Arg	Leu	Asn	Asn	Arg	Ala	Ser	Phe	Lys	Gly	Cys	Thr	Ala
				200					205					210
Leu	His	Tyr	Ala	Val	Leu	Ala	Asp	Asp	Tyr	Arg	Thr	Val	Lys	Glu
				215					220					225

Leu Leu Asp Gly	Gly Ala Asn Pro Leu Gln Arg Asn Glu Met Gly	230	235	240
His Thr Pro Leu	Asp Tyr Ala Arg Glu Gly Glu Val Met Lys Leu	245	250	255
Leu Arg Thr Ser	Glu Ala Lys Tyr Gln Glu Lys Gln Arg Lys Arg	260	265	270
Glu Ala Glu Glu	Arg Arg Arg Phe Pro Leu Glu Gln Arg Leu Lys	275	280	285
Glu His Ile Ile	Gly Gln Glu Ser Ala Ile Ala Thr Val Gly Ala	290	295	300
Ala Ile Arg Arg	Lys Glu Asn Gly Trp Tyr Asp Glu Glu His Pro	305	310	315
Leu Val Phe Leu	Phe Leu Gly Ser Ser Gly Ile Gly Lys Thr Glu	320	325	330
Leu Ala Lys Gln	Thr Ala Lys Tyr Met His Lys Asp Ala Lys Lys	335	340	345
Gly Phe Ile Arg	Leu Asp Met Ser Glu Phe Gln Glu Arg His Glu	350	355	360
Val Ala Lys Phe	Ile Gly Ser Pro Pro Gly Tyr Val Gly His Glu	365	370	375
Glu Gly Gly Gln	Leu Thr Lys Lys Leu Lys Gln Cys Pro Asn Ala	380	385	390
Val Val Leu Phe	Asp Glu Val Asp Lys Ala His Pro Asp Val Leu	395	400	405
Thr Ile Met Leu	Gln Leu Phe Asp Glu Gly Arg Leu Thr Asp Gly	410	415	420
Lys Gly Lys Thr	Ile Asp Cys Lys Asp Ala Ile Phe Ile Met Thr	425	430	435
Ser Asn Val Ala	Ser Asp Glu Ile Ala Gln His Ala Leu Gln Leu	440	445	450
Arg Gln Glu Ala	Leu Glu Met Ser Arg Asn Arg Ile Ala Glu Asn	455	460	465
Leu Gly Asp Val	Gln Ile Ser Asp Lys Ile Thr Ile Ser Lys Asn	470	475	480
Phe Lys Glu Asn	Val Ile Arg Pro Ile Leu Lys Ala His Phe Arg	485	490	495
Arg Asp Glu Phe	Leu Gly Arg Ile Asn Glu Ile Val Tyr Phe Leu	500	505	510
Pro Phe Cys His	Ser Glu Leu Ile Gln Leu Val Asn Lys Glu Leu	515	520	525
Asn Phe Trp Ala	Lys Arg Ala Lys Gln Arg His Asn Ile Thr Leu	530	535	540
Leu Trp Asp Arg	Glu Val Ala Asp Val Leu Val Asp Gly Tyr Asn	545	550	555
Val His Tyr Gly	Ala Arg Ser Ile Lys His Glu Val Glu Arg Arg	560	565	570
Val Val Asn Gln	Leu Ala Ala Ala Tyr Glu Gln Asp Leu Leu Pro	575	580	585
Gly Gly Cys Thr	Leu Arg Ile Thr Val Glu Asp Ser Asp Lys Gln	590	595	600
Leu Leu Lys Ser	Pro Glu Leu Pro Ser Pro Gln Ala Glu Lys Arg	605	610	615
Leu Pro Lys Leu	Arg Leu Glu Ile Ile Asp Lys Asp Ser Lys Thr	620	625	630
Arg Arg Leu Asp	Ile Arg Ala Pro Leu His Pro Glu Lys Val Cys	635	640	645
Asn Thr Ile				

<210> 13  
 <211> 672  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte Clone No: 1902576

<400> 13  
 Met Arg Ala Gly Arg Gly Ala Thr Pro Ala Arg Glu Leu Phe Arg  
 1 5 10 15  
 Asp Ala Ala Phe Pro Ala Ala Asp Ser Ser Leu Phe Cys Asp Leu  
 20 25 30  
 Ser Thr Pro Leu Ala Gln Phe Arg Glu Asp Ile Thr Trp Arg Arg  
 35 40 45  
 Pro Gln Glu Ile Cys Ala Thr Pro Arg Leu Phe Pro Asp Asp Pro  
 50 55 60  
 Arg Glu Gly Gln Val Lys Gln Gly Leu Leu Gly Asp Cys Trp Phe  
 65 70 75  
 Leu Cys Ala Cys Ala Ala Leu Gln Lys Ser Arg His Leu Leu Asp  
 80 85 90  
 Gln Val Ile Pro Pro Gly Gln Pro Ser Trp Ala Asp Gln Glu Tyr  
 95 100 105  
 Arg Gly Ser Phe Thr Cys Arg Ile Trp Gln Phe Gly Arg Trp Val  
 110 115 120  
 Glu Val Thr Thr Asp Asp Arg Leu Pro Cys Leu Ala Gly Arg Leu  
 125 130 135  
 Cys Phe Ser Arg Cys Gln Arg Glu Asp Val Phe Trp Leu Pro Leu  
 140 145 150  
 Leu Glu Lys Val Tyr Ala Lys Val His Gly Ser Tyr Glu His Leu  
 155 160 165  
 Trp Ala Gly Gln Val Ala Asp Ala Leu Val Asp Leu Thr Gly Gly  
 170 175 180  
 Leu Ala Glu Arg Trp Asn Leu Lys Gly Val Ala Gly Ser Gly Gly  
 185 190 195  
 Gln Gln Asp Arg Pro Gly Arg Trp Glu His Arg Thr Cys Arg Gln  
 200 205 210  
 Leu Leu His Leu Lys Asp Gln Cys Leu Ile Ser Cys Cys Val Leu  
 215 220 225  
 Ser Pro Arg Ala Gly Ala Arg Glu Leu Gly Glu Phe His Ala Phe  
 230 235 240  
 Ile Val Ser Asp Leu Arg Glu Leu Gln Gly Gln Ala Gly Gln Cys  
 245 250 255  
 Ile Leu Leu Leu Arg Ile Gln Asn Pro Trp Gly Arg Arg Cys Trp  
 260 265 270  
 Gln Gly Leu Trp Arg Glu Gly Gly Glu Gly Trp Ser Gln Val Asp  
 275 280 285  
 Ala Ala Val Ala Ser Glu Leu Leu Ser Gln Leu Gln Glu Gly Glu  
 290 295 300  
 Phe Trp Val Glu Glu Glu Glu Phe Leu Arg Glu Phe Asp Glu Leu  
 305 310 315  
 Thr Val Gly Tyr Pro Val Thr Glu Ala Gly His Leu Gln Ser Leu  
 320 325 330  
 Tyr Thr Glu Arg Leu Leu Cys His Thr Arg Ala Leu Pro Gly Ala  
 335 340 345

Trp	Val	Lys	Gly	Gln	Ser	Ala	Gly	Gly	Cys	Arg	Asn	Asn	Ser	Gly	350	355	360
Phe	Pro	Ser	Asn	Pro	Lys	Phe	Trp	Leu	Arg	Val	Ser	Glu	Pro	Ser	365	370	375
Glu	Val	Tyr	Ile	Ala	Val	Leu	Gln	Arg	Ser	Arg	Leu	His	Ala	Ala	380	385	390
Asp	Trp	Ala	Gly	Arg	Ala	Arg	Ala	Leu	Val	Gly	Asp	Ser	His	Thr	395	400	405
Ser	Trp	Ser	Pro	Ala	Ser	Ile	Pro	Gly	Lys	His	Tyr	Gln	Ala	Val	410	415	420
Gly	Leu	His	Leu	Trp	Lys	Val	Glu	Lys	Arg	Arg	Val	Asn	Leu	Pro	425	430	435
Arg	Val	Leu	Ser	Met	Pro	Pro	Val	Ala	Gly	Thr	Ala	Cys	His	Ala	440	445	450
Tyr	Asp	Arg	Glu	Val	His	Leu	Arg	Cys	Glu	Leu	Ser	Pro	Gly	Tyr	455	460	465
Tyr	Leu	Ala	Val	Pro	Ser	Thr	Phe	Leu	Lys	Asp	Ala	Pro	Gly	Glu	470	475	480
Phe	Leu	Leu	Arg	Val	Phe	Ser	Thr	Gly	Arg	Val	Ser	Leu	Ser	Ala	485	490	495
Ile	Arg	Ala	Val	Ala	Lys	Asn	Thr	Ala	Pro	Gly	Ala	Ala	Leu	Pro	500	505	510
Ala	Gly	Glu	Trp	Gly	Thr	Val	Gln	Leu	Arg	Gly	Ser	Trp	Arg	Val	515	520	525
Gly	Gln	Thr	Ala	Gly	Gly	Ser	Arg	Asn	Phe	Ala	Ser	Tyr	Pro	Thr	530	535	540
Asn	Pro	Cys	Phe	Pro	Phe	Ser	Val	Pro	Glu	Gly	Pro	Gly	Pro	Arg	545	550	555
Cys	Val	Arg	Ile	Thr	Leu	His	Gln	His	Cys	Arg	Pro	Ser	Asp	Thr	560	565	570
Glu	Phe	His	Pro	Ile	Gly	Phe	His	Ile	Phe	Gln	Val	Pro	Glu	Gly	575	580	585
Gly	Arg	Ser	Gln	Asp	Ala	Pro	Pro	Leu	Leu	Leu	Gln	Glu	Pro	Leu	590	595	600
Leu	Ser	Cys	Val	Pro	His	Arg	Tyr	Ala	Gln	Glu	Val	Ser	Arg	Leu	605	610	615
Cys	Leu	Leu	Pro	Ala	Gly	Thr	Tyr	Lys	Val	Val	Pro	Ser	Thr	Tyr	620	625	630
Leu	Pro	Asp	Thr	Glu	Gly	Ala	Phe	Thr	Val	Thr	Ile	Ala	Thr	Arg	635	640	645
Ile	Asp	Arg	Pro	Ser	Ile	His	Ser	Gln	Glu	Met	Leu	Gly	Gln	Phe	650	655	660
Leu	Gln	Glu	Val	Ser	Val	Met	Ala	Val	Met	Lys	Thr				665	670	

&lt;210&gt; 14

&lt;211&gt; 80

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2024210

&lt;400&gt; 14

Met Lys Leu Ser Gly Met Phe Leu Leu Leu Ser Leu Ala Leu Phe

1	5	10	15
Cys Phe Leu Thr Gly Val Phe Ser Gln Gly Gly Gln Val Asp Cys			
	20	25	30
Gly Glu Phe Gln Asp Pro Lys Val Tyr Cys Thr Arg Glu Ser Asn			
	35	40	45
Pro His Cys Gly Ser Asp Gly Gln Thr Tyr Gly Asn Lys Cys Ala			
	50	55	60
Phe Cys Lys Ala Ile Val Lys Ser Gly Gly Lys Ile Ser Leu Lys			
	65	70	75
His Pro Gly Lys Cys			
	80		

&lt;210&gt; 15

&lt;211&gt; 795

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2523109

&lt;400&gt; 15

Met Ala Val Leu Leu Leu Leu Leu Arg Ala Leu Arg Arg Gly Pro			
1	5	10	15
Gly Pro Gly Pro Arg Pro Leu Trp Gly Pro Gly Pro Ala Trp Ser			
	20	25	30
Pro Gly Phe Pro Ala Arg Pro Gly Arg Gly Arg Pro Tyr Met Ala			
	35	40	45
Ser Arg Pro Pro Gly Asp Leu Ala Glu Ala Gly Gly Arg Ala Leu			
	50	55	60
Gln Ser Leu Gln Leu Arg Leu Leu Thr Pro Thr Phe Glu Gly Ile			
	65	70	75
Asn Gly Leu Leu Leu Lys Gln His Leu Val Gln Asn Pro Val Arg			
	80	85	90
Leu Trp Gln Leu Leu Gly Gly Thr Phe Tyr Phe Asn Thr Ser Arg			
	95	100	105
Leu Lys Gln Lys Asn Lys Glu Lys Asp Lys Ser Lys Gly Lys Ala			
	110	115	120
Pro Glu Glu Asp Glu Glu Glu Arg Arg Arg Arg Glu Arg Asp Asp			
	125	130	135
Gln Met Tyr Arg Glu Arg Leu Arg Thr Leu Leu Val Ile Ala Val			
	140	145	150
Val Met Ser Leu Leu Asn Ala Leu Ser Thr Ser Gly Gly Ser Ile			
	155	160	165
Ser Trp Asn Asp Phe Val His Glu Met Leu Ala Lys Gly Glu Val			
	170	175	180
Gln Arg Val Gln Val Val Pro Glu Ser Asp Val Val Glu Val Tyr			
	185	190	195
Leu His Pro Gly Ala Val Val Phe Gly Arg Pro Arg Leu Ala Leu			
	200	205	210
Met Tyr Arg Met Gln Val Ala Asn Ile Asp Lys Phe Glu Glu Lys			
	215	220	225
Leu Arg Ala Ala Glu Asp Glu Leu Asn Ile Glu Ala Lys Asp Arg			
	230	235	240
Ile Pro Val Ser Tyr Lys Arg Thr Gly Phe Phe Gly Asn Ala Leu			
	245	250	255

Tyr Ser Val Gly Met Thr Ala Val Gly Leu Ala Ile Leu Trp Tyr	260	265	270
Val Phe Arg Leu Ala Gly Met Thr Gly Arg Glu Gly Gly Phe Ser	275	280	285
Ala Phe Asn Gln Leu Lys Met Ala Arg Phe Thr Ile Val Asp Gly	290	295	300
Lys Met Gly Lys Gly Val Ser Phe Lys Asp Val Ala Gly Met His	305	310	315
Glu Ala Lys Leu Glu Val Arg Glu Phe Val Asp Tyr Leu Lys Ser	320	325	330
Pro Lys Arg Phe Leu Gln Leu Gly Ala Lys Val Pro Lys Gly Ala	335	340	345
Leu Leu Leu Gly Pro Pro Gly Cys Gly Lys Thr Leu Leu Ala Lys	350	355	360
Ala Val Ala Thr Glu Ala Gln Val Pro Phe Leu Ala Met Ala Gly	365	370	375
Pro Glu Phe Val Glu Val Ile Gly Gly Leu Gly Ala Ala Arg Val	380	385	390
Arg Ser Leu Phe Lys Glu Ala Arg Ala Arg Ala Pro Cys Ile Val	395	400	405
Tyr Ile Asp Glu Ile Asp Ala Val Gly Lys Lys Arg Ser Thr Thr	410	415	420
Met Ser Gly Phe Ser Asn Thr Glu Glu Glu Gln Thr Leu Asn Gln	425	430	435
Leu Leu Val Glu Met Asp Gly Met Gly Thr Thr Asp His Val Ile	440	445	450
Val Leu Ala Ser Thr Asn Arg Ala Asp Ile Leu Asp Gly Ala Leu	455	460	465
Met Arg Pro Gly Arg Leu Asp Arg His Val Phe Ile Asp Leu Pro	470	475	480
Thr Leu Gln Glu Arg Arg Glu Ile Phe Glu Gln His Leu Lys Ser	485	490	495
Leu Lys Leu Thr Gln Ser Ser Thr Phe Tyr Ser Gln Arg Leu Ala	500	505	510
Glu Leu Thr Pro Gly Phe Ser Gly Ala Asp Ile Ala Asn Ile Cys	515	520	525
Asn Glu Ala Ala Leu His Ala Ala Arg Glu Gly His Thr Ser Val	530	535	540
His Thr Leu Asn Phe Glu Tyr Ala Val Glu Arg Val Leu Ala Gly	545	550	555
Thr Ala Lys Lys Ser Lys Ile Leu Ser Lys Glu Glu Gln Lys Val	560	565	570
Val Ala Phe His Glu Ser Gly His Ala Leu Val Gly Trp Met Leu	575	580	585
Glu His Thr Glu Ala Val Met Lys Val Ser Ile Thr Pro Arg Thr	590	595	600
Asn Ala Ala Leu Gly Phe Ala Gln Met Leu Pro Arg Asp Gln His	605	610	615
Leu Phe Thr Lys Glu Gln Leu Phe Glu Arg Met Cys Met Ala Leu	620	625	630
Gly Gly Arg Ala Ser Glu Ala Leu Ser Phe Asn Glu Val Thr Ser	635	640	645
Gly Ala Gln Asp Asp Leu Arg Lys Val Thr Arg Ile Ala Tyr Ser	650	655	660
Met Val Lys Gln Phe Gly Met Ala Pro Gly Ile Gly Pro Ile Ser	665	670	675
Phe Pro Glu Ala Gln Glu Gly Leu Met Gly Ile Gly Arg Arg Pro			



	680		685		690
Phe Ser Gln Gly	Leu Gln Gln Met Met	Asp His Glu Ala Arg	Leu		
	695		700		705
Leu Val Ala Lys	Ala Tyr Arg His Thr	Glu Lys Val Leu Gln	Asp		
	710		715		720
Asn Leu Asp Lys	Leu Gln Ala Leu Ala	Asn Ala Leu Leu Glu	Lys		
	725		730		735
Glu Val Ile Asn	Tyr Glu Asp Ile Glu	Ala Leu Ile Gly	Pro Pro		
	740		745		750
Pro His Gly Pro	Lys Lys Met Ile Ala	Pro Gln Arg Trp	Ile Asp		
	755		760		765
Ala Gln Arg Glu	Lys Gln Asp Leu Gly	Glu Glu Glu Thr	Glu Glu		
	770		775		780
Thr Gln Gln Pro	Pro Leu Gly Gly Glu	Glu Pro Thr Trp	Pro Lys		
	785		790		795

&lt;210&gt; 16

&lt;211&gt; 193

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2588566

&lt;400&gt; 16

Met Pro Asp Ser	Asp Arg His Leu Ser	Ser His Phe Asn Leu Arg
1	5	10 15
Met Lys Gly Ser	Pro Ser Glu His Gly	Ser Gln Gln Ser Ile Phe
	20	25 30
Asn Arg Tyr Ala	Gln Gln Arg Leu Asp	Ile Asp Ala Thr Gln Leu
	35	40 45
Gln Gly Leu Leu	Asn Gln Glu Leu Leu	Thr Gly Pro Pro Gly Asp
	50	55 60
Met Phe Ser Leu	Asp Glu Cys Arg Ser	Leu Val Ala Leu Met Glu
	65	70 75
Leu Lys Val Asn	Gly Arg Leu Asp Gln	Glu Phe Ala Arg Leu
	80	85 90
Trp Lys Arg Leu	Val His Tyr Gln His	Val Phe Gln Lys Val Gln
	95	100 105
Thr Ser Pro Gly	Val Leu Leu Ser Ser	Asp Leu Trp Lys Ala Ile
	110	115 120
Glu Asn Thr Asp	Phe Leu Arg Gly Ile	Phe Ile Ser Arg Glu Leu
	125	130 135
Leu His Leu Val	Thr Leu Arg Tyr Ser	Asp Ser Val Gly Arg Val
	140	145 150
Ser Phe Pro Ser	Leu Val Cys Phe Leu	Met Arg Leu Glu Ala Met
	155	160 165
Ala Lys Thr Phe	Arg Asn Leu Ser Lys	Asp Gly Lys Gly Leu Tyr
	170	175 180
Leu Thr Glu Met	Glu Trp Met Ser Leu	Val Met Tyr Asn
	185	190

&lt;210&gt; 17

&lt;211&gt; 663

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2740570

&lt;400&gt; 17

```

Met Asp Leu Leu His Glu Glu Leu Lys Glu Gln Val Met Glu Val
 1          5          10          15
Glu Glu Asp Pro Gln Thr Ile Thr Thr Glu Glu Thr Met Glu Glu
          20          25          30
Asp Lys Ser Gln Ser Asp Val Asp Phe Gln Ser Cys Glu Ser Cys
          35          40          45
Ser Asn Ser Asp Arg Ala Glu Asn Glu Asn Gly Ser Arg Cys Phe
          50          55          60
Ser Glu Asp Asn Asn Glu Thr Thr Met Leu Ile Gln Asp Asp Glu
          65          70          75
Asn Asn Ser Glu Met Ser Lys Asp Trp Gln Lys Glu Lys Met Cys
          80          85          90
Asn Lys Ile Asn Lys Val Asn Ser Glu Gly Glu Phe Asp Lys Asp
          95          100          105
Arg Asp Ser Ile Ser Glu Thr Val Asp Leu Asn Asn Gln Glu Thr
          110          115          120
Val Lys Val Gln Ile His Ser Arg Ala Ser Glu Tyr Ile Thr Asp
          125          130          135
Val His Ser Asn Asp Leu Ser Thr Pro Gln Ile Leu Pro Ser Asn
          140          145          150
Glu Gly Val Asn Pro Arg Leu Ser Ala Ser Pro Pro Lys Ser Gly
          155          160          165
Asn Leu Trp Pro Gly Leu Ala Pro Pro His Lys Lys Ala Gln Ser
          170          175          180
Ala Ser Pro Lys Arg Lys Lys Gln His Lys Lys Tyr Arg Ser Val
          185          190          195
Ile Ser Asp Ile Phe Asp Gly Thr Ile Ile Ser Ser Val Gln Cys
          200          205          210
Leu Thr Cys Asp Arg Val Ser Val Thr Leu Glu Thr Phe Gln Asp
          215          220          225
Leu Ser Leu Pro Ile Pro Gly Lys Glu Asp Leu Ala Lys Leu His
          230          235          240
Ser Ser Ser His Pro Thr Ser Ile Val Lys Ala Gly Ser Cys Gly
          245          250          255
Glu Ala Tyr Ala Pro Gln Gly Trp Ile Ala Phe Phe Met Glu Tyr
          260          265          270
Val Lys Arg Phe Val Val Ser Cys Val Pro Ser Trp Phe Trp Gly
          275          280          285
Pro Val Val Thr Leu Gln Asp Cys Leu Ala Ala Phe Phe Ala Arg
          290          295          300
Asp Glu Leu Lys Gly Asp Asn Met Tyr Ser Cys Glu Lys Cys Lys
          305          310          315
Lys Leu Arg Asn Gly Val Lys Phe Cys Lys Val Gln Asn Phe Pro
          320          325          330
Glu Ile Leu Cys Ile His Leu Lys Arg Phe Arg His Glu Leu Met
          335          340          345
Phe Ser Thr Lys Ile Ser Thr His Val Ser Phe Pro Leu Glu Gly
          350          355          360
Leu Asp Leu Gln Pro Phe Leu Ala Lys Asp Ser Pro Ala Gln Ile

```

Val Thr Tyr Asp	365	370	375
Leu Leu Ser Val Ile Cys His His Gly Thr Ala			
380	385	390	
Ser Ser Gly His Tyr Ile Ala Tyr Cys Arg Asn Asn Leu Asn Asn			
395	400	405	
Leu Trp Tyr Glu Phe Asp Asp Gln Ser Val Thr Glu Val Ser Glu			
410	415	420	
Ser Thr Val Gln Asn Ala Glu Ala Tyr Val Leu Phe Tyr Arg Lys			
425	430	435	
Ser Ser Glu Glu Ala Gln Lys Glu Arg Arg Arg Ile Ser Asn Leu			
440	445	450	
Leu Asn Ile Met Glu Pro Ser Leu Leu Gln Phe Tyr Ile Ser Arg			
455	460	465	
Gln Trp Leu Asn Lys Phe Lys Thr Phe Ala Glu Pro Gly Pro Ile			
470	475	480	
Ser Asn Asn Asp Phe Leu Cys Ile His Gly Gly Val Pro Pro Arg			
485	490	495	
Lys Ala Gly Tyr Ile Glu Asp Leu Val Leu Met Leu Pro Gln Asn			
500	505	510	
Ile Trp Asp Asn Leu Tyr Ser Arg Tyr Gly Gly Gly Pro Ala Val			
515	520	525	
Asn His Leu Tyr Ile Cys His Thr Cys Gln Ile Glu Ala Glu Lys			
530	535	540	
Ile Glu Lys Arg Arg Lys Thr Glu Leu Glu Ile Phe Ile Arg Leu			
545	550	555	
Asn Arg Ala Phe Gln Lys Glu Asp Ser Pro Ala Thr Phe Tyr Cys			
560	565	570	
Ile Ser Met Gln Trp Phe Arg Glu Trp Glu Ser Phe Val Lys Gly			
575	580	585	
Lys Asp Gly Asp Pro Pro Gly Pro Ile Asp Asn Thr Lys Ile Ala			
590	595	600	
Val Thr Lys Cys Gly Asn Val Met Leu Arg Gln Gly Ala Asp Ser			
605	610	615	
Gly Gln Ile Ser Glu Glu Thr Trp Asn Phe Leu Gln Ser Ile Tyr			
620	625	630	
Gly Gly Gly Pro Glu Val Ile Leu Arg Pro Pro Val Val His Val			
635	640	645	
Asp Pro Asp Ile Leu Gln Ala Glu Glu Lys Ile Glu Val Glu Thr			
650	655	660	
Arg Ser Leu			

&lt;210&gt; 18

&lt;211&gt; 362

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2820384

&lt;400&gt; 18

Met Tyr Ser Cys Glu Arg Cys Lys Lys Leu Arg Asn Gly Val Lys			
1	5	10	15
Tyr Cys Lys Val Leu Arg Leu Pro Glu Ile Leu Cys Ile His Leu			
20	25	30	
Lys Arg Phe Arg His Glu Val Met Tyr Ser Phe Lys Ile Asn Ser			

	35		40		45
His Val Ser Phe	Pro Leu Glu Gly Leu Asp	Leu Arg Pro Phe Leu			
	50		55		60
Ala Lys Glu Cys Thr	Ser Gln Ile Thr Thr Tyr Asp	Leu Leu Ser			
	65		70		75
Val Ile Cys His His	Gly Thr Ala Gly Ser Gly His Tyr Ile Ala				
	80		85		90
Tyr Cys Gln Asn Val	Ile Asn Gly Gln Trp Tyr Glu Phe Asp Asp				
	95		100		105
Gln Tyr Val Thr Glu	Val His Glu Thr Val Val Gln Asn Ala Glu				
	110		115		120
Gly Tyr Val Leu Phe	Tyr Arg Lys Ser Ser Glu Glu Ala Met Arg				
	125		130		135
Glu Arg Gln Gln Val	Val Ser Leu Ala Ala Met Arg Glu Pro Ser				
	140		145		150
Leu Leu Arg Phe Tyr	Val Ser Arg Glu Trp Leu Asn Lys Phe Asn				
	155		160		165
Thr Phe Ala Glu Pro	Gly Pro Ile Thr Asn Gln Thr Phe Leu Cys				
	170		175		180
Ser His Gly Gly Ile	Pro Pro His Lys Tyr His Tyr Ile Asp Asp				
	185		190		195
Leu Val Val Ile Leu	Pro Gln Asn Val Trp Glu His Leu Tyr Asn				
	200		205		210
Arg Phe Gly Gly Gly	Pro Ala Val Asn His Leu Tyr Val Cys Ser				
	215		220		225
Ile Cys Gln Val Glu	Ile Glu Ala Leu Ala Lys Arg Arg Arg Ile				
	230		235		240
Glu Ile Asp Thr Phe	Ile Lys Leu Asn Lys Ala Phe Gln Ala Glu				
	245		250		255
Glu Ser Pro Gly Val	Ile Tyr Cys Ile Ser Met Gln Trp Phe Arg				
	260		265		270
Glu Trp Glu Ala Phe	Val Lys Gly Lys Asp Asn Glu Pro Pro Gly				
	275		280		285
Pro Ile Asp Asn Ser	Arg Ile Ala Gln Val Lys Gly Ser Gly His				
	290		295		300
Val Gln Leu Lys Gln	Gly Ala Asp Tyr Gly Gln Ile Ser Glu Glu				
	305		310		315
Thr Trp Thr Tyr Leu	Asn Ser Leu Tyr Gly Gly Gly Pro Glu Ile				
	320		325		330
Ala Ile Arg Gln Ser	Val Ala Gln Arg Trp Ala Gln Arg Thr Cys				
	335		340		345
Thr Gly Ser Arg Arg	Ser Lys Pro Arg Arg Gly Pro Cys Asp Leu				
	350		355		360
Leu Gly					

&lt;210&gt; 19

&lt;211&gt; 210

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2990692

&lt;400&gt; 19

Met Val Ser Leu Leu Pro Gly Glu Pro Pro Gln Lys Ile Pro Arg

1	5	10	15
Gly Val Tyr Gly	Pro Leu Pro Glu Gly	Val Gly Leu Ile Leu	
	20	25	30
Gly Arg Ser Ser	Leu Asn Leu Lys Gly	Val Gln Ile His Thr Gly	
	35	40	45
Val Ile Tyr Ser	Asp Tyr Lys Gly Gly	Ile Gln Leu Val Ile Ser	
	50	55	60
Ser Thr Val Pro	Trp Ser Ala Asn Pro	Gly Asp Arg Ile Ala Gln	
	65	70	75
Leu Leu Leu Leu	Pro Tyr Val Lys Ile	Gly Glu Asn Lys Thr Glu	
	80	85	90
Arg Thr Gly Gly	Phe Gly Ser Thr Asn	Pro Ala Gly Lys Ala Thr	
	95	100	105
Tyr Trp Ala Asn	Gln Val Ser Glu Asp	Arg Pro Val Cys Thr Val	
	110	115	120
Thr Ile Pro Gly	Lys Glu Phe Glu Gly	Leu Val Asp Thr Gln Ala	
	125	130	135
Asp Val Ser Ile	Ile Gly Ile Gly Thr	Ala Ser Glu Val Tyr Gln	
	140	145	150
Ser Ala Met Ile	Leu His Cys Leu Gly	Ser Asp Asn Gln Glu Ser	
	155	160	165
Thr Val Gln Pro	Met Ile Thr Ser Ile	Pro Ile Asn Leu Trp Gly	
	170	175	180
Arg Asp Leu Leu	Gln Gln Trp His Ala	Glu Ile Thr Ile Pro Ala	
	185	190	195
Ser Leu Tyr Ser	Pro Arg Asn Gln Lys	Ile Met Thr Lys Met Gly	
	200	205	210

&lt;210&gt; 20

&lt;211&gt; 283

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 4590384

&lt;400&gt; 20

Met Gly Leu Gly	Leu Arg Gly Trp Gly	Arg Pro Leu Leu Thr Val
1	5	10
Ala Thr Ala Leu	Met Leu Pro Val Lys	Pro Pro Ala Gly Ser Trp
	20	25
Gly Ala Gln Ile	Ile Gly Gly His Glu	Val Thr Pro His Ser Arg
	35	40
Pro Tyr Met Ala	Ser Val Arg Phe Gly	Gly Gln His His Cys Gly
	50	55
Gly Phe Leu Leu	Arg Ala Arg Trp Val	Val Ser Ala Ala His Cys
	65	70
Phe Ser His Arg	Asp Leu Arg Thr Gly	Leu Val Val Leu Gly Ala
	80	85
His Val Leu Ser	Thr Ala Glu Pro Thr	Gln Gln Val Phe Gly Ile
	95	100
Asp Ala Leu Thr	Thr His Pro Asp Tyr	His Pro Met Thr His Ala
	110	115
Asn Asp Ile Cys	Leu Leu Arg Leu Asn	Gly Ser Ala Val Leu Gly
	125	130
		135

```

Pro Ala Val Gly Leu Leu Arg Leu Pro Gly Arg Arg Ala Arg Pro
      140      145      150
Pro Thr Ala Gly Thr Arg Cys Arg Val Ala Gly Trp Gly Phe Val
      155      160      165
Ser Asp Phe Glu Glu Leu Pro Pro Gly Leu Met Glu Ala Lys Val
      170      175      180
Arg Val Leu Asp Pro Asp Val Cys Asn Ser Ser Trp Lys Gly His
      185      190      195
Leu Thr Leu Thr Met Leu Cys Thr Arg Ser Gly Asp Ser His Arg
      200      205      210
Arg Gly Phe Cys Ser Ala Asp Ser Gly Gly Pro Leu Val Cys Arg
      215      220      225
Asn Arg Ala His Gly Leu Val Ser Phe Ser Gly Leu Trp Cys Gly
      230      235      240
Asp Pro Lys Thr Pro Asp Val Tyr Thr Gln Val Ser Ala Phe Val
      245      250      255
Ala Trp Ile Trp Asp Val Val Arg Arg Ser Ser Pro Gln Pro Gly
      260      265      270
Pro Leu Pro Gly Thr Thr Arg Pro Pro Gly Glu Ala Ala
      275      280

```

<210> 21  
 <211> 896  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> 660  
 <223> a or g or c or t, unknown, or other

<220>  
 <221> misc\_feature  
 <223> Incyte Clone No: 1220330

```

<400> 21
atgccatcgc gccggcgcca tggcatcaaa gtcattgcaga gggttcgcggg gctgccggag 60
accggccgca tggacccagg gacagtggcc accatgcgta agccccgctg ctccctgcct 120
gacgtgctgg gcgtggcggg gctggtcagg cggcgctcgc ggtacgctct gagcggcagc 180
gtgtggaaga agcgaaccct gacatggagg gtacgttcct tccccagag ctccagctg 240
agccaggaga ccgtgcgggt cctcatgagc tatgccctga tggcctgggg catggagtca 300
ggcctcacat ttcatgaggt ggattcccc cagggccagg agcccgacat cctcatcgac 360
tttgcccgcg cttccacca ggacagctac cccttcgacg gggtgggggg caccctagcc 420
catgccttct tccctgggga gcaccccatc tccggggaca ctactttga cgatgaggag 480
acctggactt ttgggtcaaa ggcctctcag cagctggagc aggagctggc aggcggctca 540
ccggttgatg aggagctggg cttcagccgg ggctggcgct tgaatcctct gggtcctggc 600
agtcctgagc gcctgagctg aatacagagg gaagaggctg ggagcaaggc cgggtgctgn 660
ggcccgcagg cctgtgttct gagagtgcct gctacgagga gctctgtggt tccccaaagga 720
gatgggaggg aagacctggg ggttgggggg ttggtacagg gggtaggggc agaaggaagg 780
gggcaagaag gcttgttgaa ccaaggggaa gaattggggg gaaggggggg aattggaatg 840
gaccttcaaa ggggggttgg taaagggggg taattgggcc ttttaaggga aggggg 896

```

<210> 22  
 <211> 4906  
 <212> DNA  
 <213> Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 1342493

&lt;400&gt; 22

```

ttgggatgtg cggagtcagt gccagcccgg tcccggccaa gcgagtggtg agccccggcgc 60
ctccaacgca acaccccgcg cctcgcgcgg ctcccccgcc gtgcggatcg gagccagccg 120
gttgttgcca tggcattcgc cagctggtgg tacaagacgc atgtcagtga aaaaaccagt 180
gaatcgctt ccaaaccagg agaaaagaaa ggatcagatg agaaaaaagc agcaagcctc 240
ggcagcagtc aatcctccag aacctatgct ggtggaacag cctcggccac caaggtgtca 300
gcttctctg gtgcaaccag caagtcttcc agtatgaatc ccacagaaac caaggctgta 360
aaaacagaac ctgagaagaa gtcacagtca accaagctgt ctgtggttca tgagaaaaaa 420
tcccagaag gaaagccaaa agaacacaca gagccaaaaa gcctacccaa gcaggcatca 480
gatacaggaa gtaacgatgc tcacaataaa aaagcagttt ccagatcagc tgaacagcag 540
ccatcagaga aatcaacaga accaaagact aaaccacaag acatgatttc tgctggtgga 600
gagagtgttg ctggtatcac tgcaatatct ggcaagccgg gtgacaagaa aaaagaaaaag 660
aaatcattaa cccagctgt gccagttgaa tctaaaccgg ataaaccatc gggaaagtca 720
ggcatggatg ctgctttgga tgacttaata gatactttag gaggacctga agaaactgaa 780
gaagaaaata caacgtatac tggaccagaa gtttcagatc caatgagttc cacctacata 840
gaggaattgg gtaaaagaga agtcacaatt cctccaaaat atagggaact attggctaaa 900
aaggaaggga tcacagggcc tcctgcagac tcttcgaaac ccatagggcc agatgatgct 960
atagacgcct tgtcatctga cttcacctgt gggtcgccta cagctgctgg aaagaaaact 1020
gaaaaagagg aatctacaga agttttaaaa gctcagtcag cagggacagt cagaagtgtc 1080
gctccacccc aagagaagaa aagaaagggtg gagaaggata caatgagtga tcaagcactc 1140
gaggctctgt cggcttcact gggcaccgg caagcagaac ctgagctcga cctccgctca 1200
attaaggaa tcgatgaggc aaaagctaaa gaagaaaaac tagagaagtg tggtaggat 1260
gatgaaacaa tcccatctga gtacagatta aaaccagcca cggataaaga tggaaaacca 1320
ctattgccag agcctgaaga aaaacccaag cctcggagtg aatcagaact cattgatgaa 1380
ctttcagaag attttgaccg gtctgaatgt aaagagaaac catctaagcc aactgaaaag 1440
acagaagaat ctaaggccgc tgcctcagct cctgtgtcgg aggtgtgtg tcggacctcc 1500
atgtgtagta tacagtcagc acccctgag ccggctacct tgaagggcac agtgccagat 1560
gatgctgtag aagccttggc tgatagcctg gggaaaaagg aagcagatcc agaagatgga 1620
aaacctgtga tggataaagt caaggagaag gccaaagaag aagaccgtga aaagcttgg 1680
gaaaaagaag aaacaattcc tcctgattat agattagaag aggtcaagga taaagatgga 1740
aagccactcc tgccaaaaga gtctaaggaa cagcttccac ccagtgatga agacttcctt 1800
ctggatgctt tgtctgagga cttctctggt ccacaaaatg cttcatctct taaatttgaa 1860
gatgctaaac ttgctgtgc catctctgaa gtggtttccc aaacccagc ttcaacgacc 1920
caagctggag cccaccccg tgatacctcg cagagtgaac aagacctcga tgatgccttg 1980
gataaactct ctgacagtct aggacaaagg cagcctgacc cagatgagaa caaaccaatg 2040
gaagataaag taaaggaaaa agctaaagct gaacatagag acaagcttgg agaaagagat 2100
gacactatcc cacctgaata cagacatctc ctggatgata atggacagga caaaccagt 2160
aagccaccta caaagaaatc agaggattca aagaaacctg cagatgacca agacccatt 2220
gatgctctct caggagatct ggacagctgt ccctccacta cagaaacctc acagaacaca 2280
gcaaaggata agtgcaagaa ggctgcttcc agctccaaag cacctaagaa tggaggtaaa 2340
gcgaaggatt cagcaaagac aacagaggaa acttccaagc caaaagatga ctaaagaaat 2400
acaagttaag gtatctggta tctgcatgta aaatcttcag ctggtggatg gtgacttttg 2460
aagaacaaaa ggctttggca acagaaaaca attgttctgg gtgatttcta gaatggtttt 2520
tgttgagtct ctgaacatcc taaatattgg tttgttattc ttttccagaa agaaaatgaa 2580
tttgactggt tcactgtgt actgagtatt gataaacttt gaattttttt aattgccttc 2640
aattggggaga gaaagcttta tattgtaag aaatatattt gataaagttt cttaaagcaa 2700
caccaaaaaa acaaaaagaa agctaagtga atttttcac attctacaca cagtgcctgt 2760
aaatctcatt tgtattttca gtttgccctt aatttttttt gttagtgttt agaaaacaat 2820
gttttaaaaa ttcttcagtg ttctgatttc ttattacccc ctttctcttt gggcttttga 2880
actgtatttg atgttgcttt gggataatgt ttataagtca aacataagat attgtacatt 2940
gggcacatat ctctcttgg gctgctaata ataaattaat aacaggtaac ctggacaaac 3000
caggaagcac caaacccctt ttcagtttga actcttcttt gccaggtgtg aggacttctg 3060
catcttacag tcagcacaga acacactgag acttgaatca agtcagcaac agagcaaaat 3120

```

```

aaagggttaga taagtccttg tgtagcaa at ttcgagcata agaaataaaa tctaattaat 3180
tcttagggta ctcatctgac ttgaactctg ttggtttact gtgtagtaa actgtgcttt 3240
ctattatcta tacataaaac ctgagcagca actgtgtctt tagagctatt gccacattag 3300
cctttgcact gtatagcgtc tggctttatg gaacttaagt ttaccaaata taaaaagaaa 3360
cttctgcttt taaaaaaatt atatatatat atattaaatt tgaaacctgc atttctccca 3420
cagcaatgta agaagtaggc tctgatgtcc taccactttg aatggttttc taatatctta 3480
atgaatagtt cctgaacatt gcactgatat catcgattag aattttgata ttaatttca 3540
tctttatttc ctggtagaga atgcaggaaa agatgtcagg tacataacat aaaacagatt 3600
gggaatttat tgtttccaaa gggcatggcc ttccttagca tcagtttgaa gcttttggtta 3660
tgacttagct gacttgtggc agcggggcaa gcaaaaacaa taacactgct tataaatggc 3720
accacatctt gttaacctcc ccccaaaata ctctctgaaa gtcatgcaca tacctatggg 3780
attttacaca ccaccagctt aaaatgctat gtctctatcc atcagaaata gtcattattc 3840
tatttttaag gcagcaacaa gaaaagaaaa aacacttttc ctgagggatt tctaaccatg 3900
tatctaatacc tcccattttg ggcagtatag gtgtttgctt ttttgttttc tttttttaag 3960
aaaaaccttg aaacctttga cactgacaga tgtgtttgca aggatacggc tgcagtatta 4020
ctaatttcca tgtgtatctg gaagtatttt taaatggcat accaaaatcc agaagttaa 4080
agatgcctat aaaagtaaac aacattttatt taaaaagaac tctgaatatg cttctttttt 4140
taattagaaa tatcttcgag acttgggtgt ttgttaataa ctaataactg gagtaagcta 4200
caggatctaa agcagccctt ttacagtct agttaggaga gagaaaataa ttgcaaata 4260
ccacttagag gcaagaaca attttttatt atcaaaaagg tttctgcaca ttgttgtggc 4320
aatattgtat ctgttttaga aatgggcttt tccaaaagca acaaaagata gggttcctcag 4380
gtgacaaaaa ctgaaaatca atatttccat gtttcattaa tcaaggcata aaatacaatt 4440
aaagcaaaat attttacatt aaaatcttgg ttgtgtattt ttttaaaaga agggaaatag 4500
tttagtttgg ggtggaaatt accagtgatt tctattttta ttaaaccatt cactacaaca 4560
aataagtata aaaattccaa ttccactttt atacctattt atttgttgta gtgaatgggt 4620
taataaatgg cagatttatg tccagaagtc actctatttt gtcgtgtatt agaggaacac 4680
attttgacat ttttcgtatc aatcatcaat catattccct tatcttgaag tttttgcctt 4740
cttatattca aaaagttcag ttgaattct cctttgccag gtgtgaggat ttccgcacct 4800
tagagtcagc gcaaaacacg ctgcaacttg aatcaatcaa gtcagcaaca gagcacaccg 4860
gacgcgtggg cggacgcgtg ggcggacgcg tgggcggacg cgtggg 4906

```

&lt;210&gt; 23

&lt;211&gt; 1641

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 1698270

&lt;400&gt; 23

```

gccgcctcca cggcccgccg tgcactgga gcgaagagcg gcctcctgaa ggaggggaag 60
ggacgtgggg gcggccacgg caggattaac ctccatttca gctaatacat ggagagatta 120
aagtctctcc tgattataac tggtttagag gtacagttcc ccttaaaaag attattgtgg 180
atgatgatga cagtaagata tggctcgtct atgacgcggg cccccgaagt atcaggtgtc 240
ctctcatatt cctgccccct gtcagtggaa ctgcagatgt ctttttcggg cagatttttg 300
ctctgactgg atgggggttac cgggttatcg ctttgcagta tccagtttat tgggaccatc 360
tcgagttctg tgatggattc agaaaacttt tagaccattt acaattggat aaagttcatc 420
tttttggcgc ttctttggga ggcttttttg cccagaaatt tgctgaatac actcacaatc 480
ctcctagagt ccattcccta atcctctgca attccttcag tgacacctct atcttcaacc 540
aaacttggac tgcaaacagc ttttggctga tgctgcatt tatgctcaaa aaaatagttc 600
ttggaaattt ttcatctggc ccggtggacc ctatgatggc tgatgccatt gatttcatgg 660
tagacaggct agaaagtttg ggtcagagtg aactggcttc aagacttacc ttgaattgtc 720
aaaattctta tgtggaacct cataaaattc gggacatacc tgtaactatt atggatgtgt 780
ttgatcagag tgcgctttca actgaagcta aagaagaaat gtacaagctg tatictaatg 840
cccgaagagc tcatctgaaa acaggaggca atttcccata cctgtgcaga agtgcagagg 900
tcaatcttta tgtacagata catttgctgc aattccatgg aaccaaatac gcggccattg 960

```



```

acccatcaat ggtcagtgcc gaggagcttg aggtgcagaa aggcagcctt ggcacagcc 1020
aggaggagca gtagtggtgc tctcgctgtc aatgatgagt tgaccoggtg tgttcttgta 1080
tagtcagtgg catcagcacc cgtcagccgg ccttttcctt caggttcgtc aggtcaccg 1140
gttctcactg tgtctgggaa gtaggactga tggatcatctt catgacaggc ggcactctca 1200
ctaagcctgt gtaactgttc cctctttggt tttcttagct tttgaatttg aagaagtact 1260
tttgaagact cccattttta gaaccgtgca gattttgcta ccaaaagtct tcaccactgt 1320
gttcttaagt gaatgttaat ttctgaggtt tgggactttg tgggtggtttt tttcttcttt 1380
tcttttccat tcttctttct tcttttttat gttgtttgct gtaaagtctg cacatccaga 1440
ttgcatatca ggacattggt tattttatgc tttcttggat ataaccatga tcagagtggc 1500
atggccacta cccactgtt tgctctcctg caaatcaact gcttttaatt tacacttaa 1560
caaattgttt tgagtgttag ctactgcctt tctagatatt agtcatttgg aataaaaaatt 1620
caatttcact gaaaaaaaaa a 1641

```

&lt;210&gt; 24

&lt;211&gt; 849

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2012492

&lt;400&gt; 24

```

aaaatgaatg aagaatgagg acattccacc actctccaag ttatcaagat gaagacccaa 60
gatggtggca ttcactctga aggtgcagca gctgagcatt ccaaatcggg gaaccaccag 120
aaaggctggc ctctcttcaa catgggatct tctggacttt tgagcctcct ggtgctattc 180
gtcctcttag cgaatgtcca gggacctggt ctgactgatt ggttatttcc caggagatgt 240
cccaaaatca gagaagaatg tgaattccaa gaaagggatg tgtgtacaaa ggacagacaa 300
tgccaggaca acaagaagtg ttgtgtcttc agctgcggaa aaaaatgttt agatctcaaa 360
caagatgtat gcgaaatgcc aaaagaaact ggccccctgcc tggcttattt tcttcattgg 420
tggtagtaca agaaagataa tacttgctcc atgtttgtct atggtggctg ccagggaac 480
aataacaact tccaatccaa agccaactgc ctgaacacct gcaagaataa acgctttccc 540
tgattggata aggatgcact ggaagaactg ccagaatgtg gctcatgctc tgagtactgt 600
tctgtacct gacggatgct ccagactggc ttccagtttc actctcagca ttccaagatc 660
ttagcccttc ccagaacaga acgcttgcat ctacctcttc tctctccatc tttggctctt 720
ttgatgcaca atatccatcc gttttgattt catctttatg tcccccttat ctccaacttc 780
tagaactccc agttttatacc tgtgtcactc tcaatttttt ccagtaaagt acttgatgta 840
aaaaaaaaa 849

```

&lt;210&gt; 25

&lt;211&gt; 2166

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2309875

&lt;400&gt; 25

```

ggaccaacaa agatggcggc ggccccctgcg gcgggagcga tctgggcaac ggctgcggct 60
aaagctgcag ccggggccac ggggggggctg cacgggggta gtagggggtg gccctgaact 120
ggggcctggc cctggctggc ctctcccgcc gcctcactgg gggacaggtc cagcctgttg 180
tgtccacaat gcccaggcc tctgagcacc gcctgggccc taccgagag ccacctgtta 240
atatccagcc ccgagtggga tccaagctac catttgcccc caggggccgc agcaaggagc 300
gcagaaaccc agcctctggg ccaaacccca tgttacgacc tctgcctccc cggccaggtc 360
tgctgatga acggctcaag aaactggagc tgggacgggg acggacctca ggccctcgtc 420
ccagaggccc ccttcgagca gatcatgggg tccccctgcc tggctacca cccccaacag 480

```

```

tggctttgccc tctcccatct cggaccaact tagcccggtc caagtctgtg agcagtgggg 540
acttgcgctcc aatgggggatt gccttggggag ggcaccgtgg caccggagag cttggggctg 600
cactgagccg cttggccctc cggcctgagc caccacttt gagacgtagc acttctctcc 660
gccgcctagg gggctttcct ggaccccccta cctgtttcag catacggaca gagccccctg 720
cttcccatgg ctcttccac atgatattccg cccggtcctc tgagcctttc tactctgatg 780
acaagatggc tcatcacaca ctcttctgtg gctctgggtca tgttggcctt cgaaacctgg 840
gaaacacgtg ctctctgaat gctgtgctgc agtgtctgag cagcactcga cctcttcggg 900
acttctgtct gagaaggac ttccggcaag aggtgcctgg aggaggccga gcccaagagc 960
tactgaagc ctttgagat gtgattgggt cctctggga cctgactcc tgcgaagctg 1020
tgaatcctac tcatccga gctgtcttc agaatatgt tccctcctc tctggataca 1080
gccagcagga tgcccaagag ttctgaagc tctcatgga gcggctacac cttgaaatca 1140
accgcccagg ccgcccgggt ccaccgatac ttgccaatgg tccagttccc tctccacccc 1200
gccgaggagg ggctctgcta gaagaacctg agttaagtga tgatgaccga gccaacctaa 1260
tgtggaaacg ttacctggag cgagaggaca gcaagattgt ggacctgtt gtgggcccagt 1320
tgaaaagtgt tctcaagtgc caggcctgtg ggtatcgctc cagcacttc gaggtttttt 1380
gtgacctgtc cctgcccac cccaagaaag gatttgctgg gggcaagggt tctctgcggg 1440
attgtttcaa ctttttact aaggaagaag agctagagtc ggagaatgcc ccagtgtgtg 1500
accgatgtcg gcagaaaact cgaagtacca aaaagttgac agtacaaga ttccctcgaa 1560
tctctgtgct ccatctgaat cgattttctg cctcccagg ctccatcaaa aaaagttcag 1620
taggtgtaga ctttccactg cagcgactga gcttagggga ctttgccagt gacaaagccg 1680
gaagtctgt ataccagctg tatgccctt gcaaccactc aggcagcgtc cactatggcc 1740
actacacagc cctgtgccg tgccagactg gttggcatgt ctacaatgac tctctgtgtc 1800
ccctgtcag tgaaaaccag gtggcatcca gcgagggcta cgtgtgttc taccaactga 1860
tgcaggagcc accccggtgc ctgtgacacc tctaagctct ggcacctgtg aagcccttta 1920
aacaccctta agccccaggc tccccgttta cctcagagac gtctattttt gtgtcttttt 1980
aatcggggag gggggagggg gtggtttag ctccattatt ttttttatta aaaaatcccc 2040
ttccacctgg aggtccctt gtctcccag cccatgtaca aagctcacca agccctgcc 2100
catgtacagc cccagaccc tctgcaatat cactttttgt gaataaattt attaagaaaa 2160
aaaaaa

```

&lt;210&gt; 26

&lt;211&gt; 2069

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2479394

&lt;400&gt; 26

```

gacttgagtc actctcagac tctttataaa tacagcttga ctcagccact gtatgactga 60
ctccccgggg acatgagggt gatactgttc attggggccc ttattgggtc cagcatctgt 120
ggccaagaaa aatttttttg ggaccaagtt ttgaggatta atgtcagaaa tggagacgag 180
atcagcaaat tgagtcaact agtgaattca aacaacttga agctcaattt ctggaaatct 240
cctctctcct tcaatcggcc tgtggatgtc ctggtcccat ctgtcagtct gcaggcattt 300
aaatccttcc tgagatccca gggcttagag tacgcagtga caattgagga cctgcaggcc 360
cttttagaca atgaagatga tgaaatgcaa cacaatgaag ggcaagaacg gagcagtaat 420
aacttcaact acggggctta ccattccctg gaagctattt accacgagat ggacaacatt 480
gccgcagact ttctgacct ggcgaggagg gtgaagattg gacattcgtt tgaaaaccgg 540
ccgatgtatg tactgaagtt cagcactggg aaaggcgtga ggcgccggc cgtttggctg 600
aatgcaggca tccattccc agagtggatc tcccaggcca ctgcaatctg gacggcaagg 660
aagattgtat ctgattacca gagggatcca gctatcacct ccatcttggg gaaaatggat 720
attttcttgt tgcctgtggc caatcctgat ggatattgtg atactcaaac tcaaaaccga 780
ttatggagga agacgcggtc ccgaaatcct ggaagctcct gcattgggtg tgacccaaat 840
agaaactgga acgctagttt tgcaggaaag ggagccagcg acaacccttg ctccgaagtg 900
taccatggac cccacgcaa ttccggaagt gaggtgaaat cagtggtaga tttcatccaa 960
aaacatggga atttcaaggg cttcatcgac ctgcacagct actcgcagct gctgatgtat 1020

```

```

ccatatgggt actcagtcaa aaaggcccca gatgccgagg aactcgacaa ggtggcgagg 1080
cttgcgggcca aagctctggc ttctgtgtcg ggcactgagt accaagtggg tcccacctgc 1140
accactgtct atccagctag cgggagcagc atcgactggg cgtatgacaa cggcatcaaa 1200
tttgcatcca catttgagtt gagagatacc gggacctatg gcttcctcct gccagctaac 1260
cagatcatcc ccactgcaga ggagacgtgg ctggggctga agaccatcat ggagcatgtg 1320
cgggacaaac tctactaggc gatggctctg ctctgtctac atttatattgt acccacacgt 1380
gcacgcactg aggccattgt taaaggagct ctttcctacc tgtgtgagtc agagccctct 1440
gggtttgtgg agcacacagg cctgcccctc tccagccagc tccctggagt cgtgtgtcct 1500
ggcgtgtcc ctgcaagaac tggttctgcc agcctgctca attttgggtcc tgtgtgtttt 1560
gatgagcctt ttgtctgttt ctccctccac cctgctggct gggcggtgc actcagcatc 1620
acccttcctt ggggtggcatg tctctctcta cctcattttt agaaccaaaag aacatctgag 1680
atgattctct accctcatcc acatctagcc aagccagtga ccttgctctg gtggcactgt 1740
gggagacacc acttgctctt aggtgggtct caaagatgat gtagaatttc ctttaatttc 1800
tcgcagtctt cctggaaaat attttccttt gagcagcaaa tctttagagg atatcagtga 1860
aggtctctcc ctccctcctc tctgtttttt ttttttttga gacagagttt tgctcttgtt 1920
gcccaggctg gagtttaagg gcccgatctg ggctcaccaa aacctttgcc tccgggggtc 1980
aagaaatttt cctgcctaag ccttttgagt accttgggtt aaaggggcat tcaacaatgc 2040
ctgggaaatt ttgggttttt agaagaaaa 2069

```

<210> 27

<211> 2490

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte Clone No: 2613215

<400> 27

```

ccgcgggtga tcagctggtc tgcgtccccc tgacgtgggc tggggcacgt caccgccgaa 60
tggcagcctc cagaaagcca ccgcgagtaa gggatgaatca ccaggatttt caactgagaa 120
atttaagaat aattgaacct aacgagggtga cacactcagg agacacagggt gtggaaacag 180
acggcagaat gcctccaaag gtgacttcag agctgcttcg gcagctgaga caagccatga 240
ggaactctga gtatgtgacc gaaccgatcc aggcctacat catcccatcg ggagatgctc 300
atcagagtga gtatattgct ccatgtgact gtcggcgggc ttttgtctct ggattcgatg 360
gctctgcggg cacagccatc atcacagaag agcatgcagc catgtggact gacggggcgt 420
actttctcca ggctgccaag caaatggaca gcaactggac acttatgaag atgggtctga 480
aggacacacc aactcaggaa gactggctgg tgagtgtgct tccgaagggt tccaggggtg 540
gtgtggaccc cttgatcatt cctacagatt attggaagaa aatggccaaa gttctgagaa 600
gtgccggcca tcacctcatt cctgtcaagg agaacctcgt tgacaaaatc tggacagacc 660
gtcctgagcg cccttgcaag cctctcctca cactgggcct ggattacaca ggcatctcct 720
ggaaggacaa ggttgagac cttcggttga aaatggctga gaggaacgtc atgtggtttg 780
tggtcactgc cttggatgag attgctggc tatttaatct ccgaggatca gatgtggagc 840
acaatccagt atttttctcc tacgcaatca taggactaga gacgatcatg ctcttcattg 900
atgggtgacc catagacgcc ccagtggtga aggagcacct gcttcttgac ttgggtcttg 960
aagccgaata caggatccag gtgcatccct acaagtccat cctgagcgag ctcaaggccc 1020
tgtgtgctga cctctcccca agggagaagg tgtgggtcag tgacaaggcc agctatgctg 1080
tgagcgagac catccccaag gaccaccgct gctgtatgcc ttacaccccc atctgcatcg 1140
ccaaagctgt gaagaattca gctgagtcag aaggcatgag gcgggctcac attaaagatg 1200
ctgttgctct ctgtgaactc ttttaactggc tggagaaaga ggttcccaa ggtggtgtga 1260
cagagatctc agctgctgac aaagctgagg agtttcgcag gcaacaggca gactttgtgtg 1320
acctgagctt cccaacaatt tccagtacgg gacccaacgg cgccatcatt cactacgcgc 1380
cagtccttga gacgaatagg accttgctcc tggatgaggt gtaccttatt gactcgggtg 1440
ctcaatacaa ggatggcacc acagatgtga cgcggacaat gcattttggg acccctacag 1500
cctacgagaa ggaatgcttc acatatgtcc tcaagggcca catagctgtg agtgcagccg 1560
ttttcccgac tggaaaccaa ggtcaccttc ttgactcctt tgcccgttca gctttatggg 1620
attcaggcct agattacttg cacgggactg gacatgggtg tgggtctttt ttgaatgtcc 1680

```

```

atgaggggtcc ttgcggcatc agttacaaaa cattctctga tgagcccttg gaggcaggca 1740
tgattgtcac tgatgagccc ggggtactatg aagatggggc ttttggaatt cgcattgaga 1800
atgttgtcct tgtgggtcct gtgaagacca agtataattt taataaccgg ggaagcctga 1860
cctttgaacc tctaaccattg gttccaattc agacaaaaat gatagatgtg gattctctta 1920
cagacaaaga gtgcgactgg ctcaacaatt accacctgac ctgcagggat gtgattggga 1980
aggaattgca gaaacagggc cgccaggaag ctctcgagtg gctcatcaga gagacgcaac 2040
ccatctccaa acagcattaa taaatacctc cccggttttg tttttgtaaa atgctctgga 2100
ggaaggaaga aacgtggcag atccctgaca tctttccctt ttcctttcct tcttccctac 2160
ctcccccttt tacttttagac tttaagaaga acagaaaatc ttcttatcct ctttgatatt 2220
ttattgcaaa cactcagtct tttatgattt ttttaattgtt gagaacaagc caagaataaa 2280
attgctgcac cagaaggagg gtccctccaa agttgaacac ttgggtgaaag gaagatgccc 2340
cgacttcttt ggccagtgat ggggaatcag tgagtgtccc atgatgggtca tgttccaggt 2400
gctagtacat cattcatgat caccttaatg ctcatgagac tatatttatg atcagtgaat 2460
aaaaatgtca gaactgtgaa aaaaaaaaaa

```

2490

&lt;210&gt; 28

&lt;211&gt; 3148

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 001528

&lt;400&gt; 28

```

gaagatggcg gaaggcggag cggcggatct ggacacccag cggctctgaca tcgcgacgct 60
gctcaaaaacc tcgctccgga aaggggacac ctggtaccta gtcgatagtc gctgggttcaa 120
acagtggaaa aaatatgttg gctttgacag ttgggacaaa taccagatgg gagatcaaaa 180
tgtgtatcct ggacccattg ataactctgg acttctcaaa gatggtgatg cccagtcact 240
taaggaacac cttattgatg aattggatta catactgttg ccaactgaag gttggaataa 300
acttgtcagc tggtagacat tgatggaagg tcaagagcca atagcacgaa aggtgggttga 360
acagggtatg ttgttaaagc actgcaaagt agaagtatat ctacacagaat tgaagctatg 420
tgaaaatgga aacatgaata atgttgtaac tcgaagattt agcaaagctg acacaataga 480
tacaattgaa aaggaaataa gaaaaatctt cagtattcca gatgaaaagg agaccagatt 540
gtggaacaaa tacatgagta acacatttga accactgaat aaaccagaca gcaccattca 600
ggatgctggt ttataccaag gacaggatatt agtgatagaa cagaaaaatg aagatggaaac 660
acggccaagg ggtccttcta ctctaatgt gaaaaactca aattactgtc ttccatcata 720
taccgcttat aagaactatg attattcgga acctggaaga aacaatgaac agccaggcct 780
ctgtggccta agtaacttgg gaaatacgtg tttcatgaac tcagctattc agtgtttgag 840
caacacacct ccacttactg agtatttcct caatgataag tatcaagaag aactgaattt 900
tgacaatccc ttaggaatga gaggtgaaat agctaaatct tatgccgaac tgatcaagca 960
aatgtggtct ggaaagttaa gctacgtcac cccaagagcc ttttaagacac aggtaggacg 1020
ttttgcacct cagttctctg gatatcagca gcaagactgt caagaactgt tagctttcct 1080
attagatgga ttacatgagg atttgaatag aattaggaaa aaaccatata tacaattaaa 1140
agatgcagat ggaaggccag ataagggtgt tgccgaagaa gcctgggaaa accattttaa 1200
acgaaatgat tctatcatag tagatatatt tcatggcctt ttcaaatcaa ctttagtttg 1260
tcttgagtgt gctaagattt cagtaacatt tgatcctttt tgttacttga cacttccatt 1320
gcccatgaaa aaagaacgca ccttgggaagt ttacttagtt agaatggatc cacttaccaa 1380
acctatgcag taaaagtgg ttgtcccaaa aattggaaac atattagatc tttgtacagc 1440
attgtctgct ttgtcaggaa tacctgcaga taagatgata gttactgata tatacaatca 1500
tagatttcac agaattattc ctatggatga aaaccttagt agtattatgg aacgggatga 1560
tatttatgtg ttgtgaaatta acatcaatag gacagaagat acagagcacg tgattattcc 1620
tgtttgcccta agagaaaaat tcagacactc gagttatacc caccatactg gttcttccact 1680
ttttggtcag ccctttctta tggctgtacc acgaaacaat actgaagaca aactttataa 1740
tctcctgctc ttgagaatgt gccgatatgt caaaatatct actgaaactg aagaaactga 1800
aggatcccta cactgctgta aggacaaaaa tattaatggg aatggcccaa atggcataca 1860
tgaagaaggc tcaccaagtg aaatggaaac agatgagcca gatgatgaat ccagccagga 1920

```

```

tcaagaactt ccctcagaga atgaaaacag tcagtctgaa gattcagttg gaggagataa 1980
tgattctgaa aatggattat gtactgagga tacttgcaaa ggtcaactca cgggacacaa 2040
aaaacgattg tttacattcc agttcaacaa cttaggcaat actgatatca actacatcaa 2100
agatgatacc aggcataata gatttgatga taggcagctt aggctagatg aaagatcttt 2160
tcttgctttg gattggggtc ctgatttgaa aaaaagatat tttgatgaaa atgctgctga 2220
ggactttgaa aaacatgaaa gtgtggagta taaacctcct aaaaaaccct ttgtgaaatt 2280
aaaagattgc attgaacttt ttacaacaaa agaaaagcta ggtgctgaag atccctggta 2340
ttgtccgaat tgtaaagaac atcagcaagc cacaaagaaa ttggatttat ggtccctgcc 2400
tccagtactt gtagtacatc tcaagcgatt ttcttacagt cgatacatga gagacaagtt 2460
ggatacctta gttgattttc ctatcaatga cttggatatg tcggaattct taattaatcc 2520
aaatgcaggt ccttgccgct ataactctgat tgctgtttcc aaccactatg gagggatggg 2580
aggaggacac tatactgctt ttgcaaaaaa taaagatgat ggaaaatggg actattttga 2640
tgacagtagt gtctccactg catctgaaga ccaaattgtg tccaaagcag catatgtact 2700
cttctaccag agacaagaca ctttcagtgg aactggcttt tttcctcttg accgagaaac 2760
taaaggtgct tcagctgcca ctggcatccc attagaaagt gatgaagata gcaatgataa 2820
tgacaatgat atagaaaatg aaaactgtat gcacactaac taatgaaagt cctagaagcc 2880
ataaaagaga cactttcctg ctggtggtat ctatggaaat gatgaagtta ccaccacat 2940
taaaacaaaa gtctgagatg gggagtttca gataaccgaa tgtaaatcct ttatcagatt 3000
ttaacttggt cagtacttga agtgaaacac aatgaaaact ttaacagaaa ttgtctctta 3060
atacatttac agtcttgtat ttacaagcta aatatatata ggaaatcaca aataaatccc 3120
ttttaagttt gctgctgttt tgattaaa 3148

```

<210> 29

<211> 855

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> 747

<223> a or g or c or t, unknown, or other

<220>

<221> misc\_feature

<223> Incyte Clone No: 998626

<400> 29

```

caggcaggca tcccccgag ttgtctcttt tcatgccagc gccaacagga ggctgtcttg 60
acacactgat tactactca ccagcctcct tcttttgtec accagcccc ctcttttgte 120
caccagccca gcctgactcc tggagattgt gaatagctcc atccagcctg agaaacaagc 180
cgggtggctg agccaggctg tgcacggagc gcctgacggg cccaacagac ccatgctgca 240
tccagagacc tcccctggcc gggggcatct cctggctgtg ctctggccc tccttggcac 300
cgctgggca gaggtgtggc caccacagct gcaggagcag gctccgatgg ccggagccct 360
gaacaggaag gagagtttct tgctcctctc cctgcacaac cgctgcgca gctgggtcca 420
gccccctgcg gctgacatgc ggaggctgga ctggaagtac agcctggccc aactggctca 480
agccagggca gccctctgtg gaatcccaac cccgagcctg gcgtccggcc tgtggcgcac 540
cctgcaagtg ggctggaaca tgcagctgct gcccgcgggc ttggcgtoct ttgttgaagt 600
ggtcagccta tggtttgagc aggggcagcg gtacagccac gcggcaggag agtgtgctcg 660
caacgccacc tgcaccact acacgcagct cgtgtgggac acctcaagcc agctgggctg 720
tggggcgcac tagtgtctgc agggccangag catagaagct ttgtctgtgc tactccccgg 780
aggcactggg agtcacggga gacatcatcc tataagaggg tgctgtgtcg tctgacagca 840
tgttcagctg ctcaa 855

```

<210> 30

<211> 1912

<212> DNA

<213> Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 1393301

&lt;400&gt; 30

```
agaattcggc acgaggggtta gaggcgggctt gtgtccacgg gacgcgggcg gatcttctcc 60
ggccatgagg aagccagccg ctggcttcct tccctcactc ctgaagggtgc tgctcctgcc 120
tctggcacct gccgcagccc aggatcgac tcaggcctcc actccaggca gccctctctc 180
tcctaccgaa tacgaacgct tcttcgcact gctgactcca acctggaagg cagagactac 240
ctgccgtctc cgtgcaaccc acggctgccc gaatcccaca ctgctccagc tggaccaata 300
tgaaaaccac ggcttagtgc ccgatgggtc tgtctgctcc aacctccctt atgcctcctg 360
gtttgagtct ttctgccagt tcaactacta ccgttgctcc aaccacgtct actatgccaa 420
gagagtcctg tgttcccagc cagtctctat tctctcacct aacactctca aggagataga 480
agcttcagct gaagtctcac ccaccacgat gacctcccc atctcacccc acttcacagt 540
gacagaacgc cagaccttcc agccctggcc tgagaggctc agcaacaacg tgggaagagct 600
cctacaatcc tccttgctcc tgggaggcca ggagcaagcg ccagagcaca agcaggagca 660
aggagtggag cacaggcagg agccgacaca agaacacaag caggaagagg ggcagaaaca 720
ggaagagcaa gaagaggaac aggaagagga gggaaagcag gaagaaggac aggggactaa 780
ggagggacgg gaggtgtgtg ctgagctgca gacagactca gagcccaagt ttcactctga 840
atctctatct tctaaccctt cctcttttgc tccccgggta cgagaagtag agtctactcc 900
tatgataatg gagaacatcc aggagctcat tcgatcagcc caggaaatag atgaaatgaa 960
tgaaatatat gatgagaact cctactggag aaaccaaacc cctggcagcc tcctgcagct 1020
gccccacaca gaggccttgc tgggtgctgt ctattcgatc gtggagaata cctgcatcat 1080
aacccccaca gccaaggcct ggaagtacat ggaggaggag atccttggtt tcgggaagtc 1140
ggtctgtgac agccttgggc ggcgacacat gtctacctgt gccctctgtg acttctgtc 1200
cttgaagctg gagcagtgcc actcagaggc cagcctgcag cggcaacaat gcgacacctc 1260
ccacaagact ccctttgtca gcccttgct tgcctcccag agcctgtcca tcggcaacca 1320
ggtagggtcc ccagaatcag gccgctttta cgggctggat ttgtacggtg ggctccacat 1380
ggacttctgg tgtgcccgcc ttgccacgaa aggctgtgaa gatgtccgag tctctgggtg 1440
gtccagact gagttcetta gcttccagga tggggatttc cctaccaaga tttgtgacac 1500
agactatata cagtacccaa actactgttc cttcaaaagc cagcagtgct tgatgagaaa 1560
ccgcaatcgg aaggtgtccc gcatgagatg tctgcagaat gagacttaca gtgcgctgag 1620
ccctggcaaa agtgaggacg ttgtgcttcg atggagccag gagttcagca ccttgactct 1680
aggccagttc ggtatgagctg gcgtctattc tgcccacacc ccagcccaac ctgcccacgt 1740
tctctattgt tttgagaccc cattgctttc aggtcgcccc ttctgggtct gttactcggc 1800
ccctactcac atttccttgg gttggagcaa cagtcccaga gagggccatg gtgggagctg 1860
cgccctcctt aaaagatgac tttacataaa atgttgatct tcaaaaaaaaa aa 1912
```

&lt;210&gt; 31

&lt;211&gt; 768

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 1444055

&lt;400&gt; 31

```
taagtgttac aattttaaagc caaaggcata ccaaatatta aagcaacaca acataaggaa 60
tcccattcac cacacaccat ttaaacacaa atgatggatt ctgttattta tcaaacccat 120
ttaatgtttt gtgtagtgtt ttcattggatt ccagggaatgg cttccctttc tcaaatgac 180
aatttcaagt ctattctctc atatttccca tctgattcca ctatgagcag tgccattgca 240
tgccctacta gaactgtgga tacgctttta aacaagtgga agatgatagg gtgggactca 300
ttacgtctaa tcttggggaa cactgataac gtgtccagga gagacagcac aaggggctcc 360
atcttcatca cacaactcat cgcatgcttc cagagatatt cctggcgctg ccacctagag 420
gaagtatttt ggaaggttca gcaagcattt gaaagtccgg aggcaacagt ccaaatgccc 480
accatagaac gagtgtccat gacaagatat ttctacctct ttctggcaa ctgaaaatgg 540
```

```

ttaagcattg agagtgtgtg gtggtgtatg aaataaatga aagtgtgata ttggagcagg 600
aaaccacaag cagcccagcc ctccctttatc aacttcaaga aacaccttta ctagtacaga 660
ttgaatgctt aacattttga atttcaataa aggtgaagac aaatgaaaaa aataaaaaaa 720
aaacaaaaac aaaaaacaaa caaaaaaaca aaaaaaaaat aaaaacgg 768

```

<210> 32

<211> 2069

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte Clone No: 1650177

<400> 32

```

gcacaggggc cggcaccacg ggggttatcg agcagctgtc aagatgctgg ggtccctggt 60
gttgaggaga aaagcactgg cgccacggct actcctccgg ctgctcaggt cccaacgct 120
ccggggccat ggaggtgctt ccggccggaa tgtgactact gggagtctcg gggagccgca 180
gtggctgagg gtagccaccg gggggcgccc tggaacatcg ccggccttgt tctccggacg 240
tggggcagcc accggggggc gccagggagg acgcttcgat accaaatgcc tcgcggtcgc 300
cacttgggga cgccttcctg gtcccgaaga aacactccca ggacaggaca gctggaacgg 360
gggtccccagc agggccggac tgggcatgtg cgccctggcc gcagcgtggt tggttcattg 420
ctacagcaag agtccgtcca acaaggatgc agccctgttg gaagctgccc gtgccacaa 480
tatgcaagaa gtcagcagtg tggtagaggt cctgcttgct gctggggctg atccaaacct 540
tggagatgat ttcagcagtg ttttcaagac tgccaaggaa cagggaatcc attccttggg 600
agtctgctac acccgagagg atgacttcaa caacaggctg aacaaccgcg ccagtttcaa 660
gggctgcacg gccttgcaat atgctgttct tgctgatgac taccgcaact ccaggagct 720
gcttgatgga ggagccaacc ccctgcagag gaatgaaatg ggacacacac ccttggtatta 780
tgcccgagaa ggggaagtga tgaagcttct gaggacttct gaagccaagt accaagagaa 840
gcagcgggaag cgtgaggctg aggagcggcg ccgcttcccc ctggagcagc gactaaagga 900
gcacatcatt ggccaggaga gcgccatcgc cacagtgggt gctgcgatcc ggaggaagga 960
gaatggctgg tacgatgaag aacaccctct ggtcttctct ttcttgggat catctggaat 1020
aggaaaaaca gagctggcca agcagacagc caaatatatg cacaagatg ctaaaaaggg 1080
cttcatcagg ctggacatgt ccgagttcca ggagcgacac gaggtggcca agtttattgg 1140
gtctccacca ggctacgttg gccatgagga ggggtggccag ctgaccaaga agttgaagca 1200
gtgccccaat gctgtggtgc tctttgatga agtagacaag gcccatccag atgtgctcac 1260
catcatgctg cagctgtttg atgagggccg gctgacagat ggaaaaggga agaccattga 1320
ttgcaaggac gccatcttca tcatgacctc caatgtggcc agcgacgaga tcgcacagca 1380
cgcgctgcag ctgaggcagg aagctttgga gatgagccgt aaccgtattg ccgaaaacct 1440
gggggatgtc cagataagtg acaagatcac catctcaaag aacttcaagg agaatgtgat 1500
tcgccctatc ctgaaagctc acttccggag ggatgagttt ctgggacgga tcaatgagat 1560
cgtctacttc ctccccctct gccactcgga gctcatccaa ctctgcaaca aggaactaaa 1620
cttctggggc aagagagcca agcaaaggca caacatcacg ctgctctggg accgcgaggt 1680
ggcagatgtg ctggctgacg gctacaatgt gcactatggc gcccgtcca tcaaactga 1740
ggtagaacgc cgtgtggtga accagctggc agcagcctat gagcaggacc tgctgccagg 1800
gggctgtact ttgcgcatca cgggtggagga ctgagacaag cagctactca aaagcccaga 1860
actgccctca cccagggctg agaagcgctt cccaagctg cgtctggaga tcatcgacaa 1920
ggacagcaag actcgagac tggacatccg ggcaccactg caccctgaga aggtgtgcaa 1980
caccatctag cagccacctg cctgctccta tgtgccctca ccatccaata aaggccccct 2040
ggctgtggca tggcaaaaaa aaaaaaaaaa

```

<210> 33

<211> 2594

<212> DNA

<213> Homo sapiens

<220>

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 1902576

&lt;400&gt; 33

```
ccagcacctg cggggccctc gggcttgga ggcctgggccc gacggtgaac ggtcggcgcg 60
ggccggatcg gcggcggtcg actcgcttc tctccggggc tgcgaccccg aggcaaccgg 120
ctgcagatgg gagcccgcg agcccggtat gcggcggggc cggggcgcgga cgccggcgag 180
ggagctgttc cgggacgccc ccttccccgc cgcggactcc tcgctcttct gcgacttgtc 240
tacgcccgtg gccaggttc gcgaggacat cacgtggagg cggccccagg agatttgtgc 300
cacaccccg ctgtttccag atgaccacg ggaagggcag gtgaagcagg ggctgctggg 360
ggattgctgg ttctgtgtg cctgcgcgc gctgcagaag agcaggcacc tcctggacca 420
ggtcattcct ccgggacagc cgagctgggc cgaccaggag taccggggct ccttcacctg 480
tcgcatttgg cagtttgac gctgggtgga ggtgaccaca gatgaccgcc tgccgtgcct 540
tgcagggaga ctctgtttct cccgctgcca gagggaggat gtgttctggc tccccttact 600
ggaaaagggtc tacgccaagg tccatgggtc ctacgagcac ctgtggggccg ggaggtggc 660
ggatgccttg gtggacctga ccggcgccct ggcagaaaga tggaacctga agggcgtagc 720
aggaagcgga ggccagcagg acaggccggg ccgctgggag cacaggactt gtcggcagct 780
gctccacctg aaggaccagt gtctgatcag ctgctgcgtg ctacgcccc gaagcaggtg 840
ccgggagctg ggggagttcc atgccttcat tgtctcggac ctgcgggagc tccagggtca 900
ggcggggccag tgcattcctg tctgcggat ccagaacccc tggggccggc ggtgctggca 960
ggggctctgg agagaggggg gtgaaggggt gagccaggta gatgcagcgg tagcatctga 1020
gctcctgtcc cagctccagg aaggggagtt ctgggtggag gaggaggagt tcctcaggga 1080
gtttgacgag ctaccggtg gctaccgggt cacggaggcc ggccacctgc agagcctcta 1140
cacagagagg ctgctctgcc ataccggggc gctgcctggg gcctgggtca agggcgagtc 1200
agcaggaggc tgccggaaca acagcggtt tcccagcaac cccaaattct ggctgcgggt 1260
ctcagaaccg agtgagggtg acattgcgt cctgcagaga tccaggctgc acgcgcgga 1320
ctgggcaggc cggggccggg cactggtggg tgacagtcac acttcgtgga gccagcgag 1380
catcccgggc aagcactacc aggtgtgtgg tctgcacctc tggaaggtag agaagcgcg 1440
ggtcaatctg cctaggggtc tgtccatgcc ccccggtggc ggcaccgcgt gccatgcata 1500
cgaccgggag gtccacctgc gttgtgagct ctaccgggc tactacctgg ctgtccccag 1560
caccttctg aaggacgcgc caggggagtt cctgctccga gtcttctcta cgggcgagt 1620
ctcccttagc gccatcaggg cagtggccaa gaacaccgcc cccggggcag ccctgcctgc 1680
gggggagtg gggaccgtgc agctacggg ttcttgga gtcggccaga cggcgggggg 1740
cagcaggaac tttgcctcat accccaccaa cccctgcttc cccttctcgg tccccgaggg 1800
ccctggcccc cgtgcgtcc gcatactct gcatacgac tgccggccca gtgacaccga 1860
gttccacccc atcggttcc atatcttcca ggtcccagag ggtggaagga gccaggacgc 1920
acccccactg ctgctgcagg agccgtgct gagctgcgtg ccacatcgct acgcccagga 1980
ggtgagccgg ctctgcctcc tgcctgcggg cacctacaag gttgtgccct ccacctacct 2040
gccggacaca gagggggcct tcacagtgc catcgcaacc aggattgaca ggccatccat 2100
tcacagccag gagatgctgg gccagttcct ccaagaggtc tccgtcatgg cagtgatgaa 2160
aacctaacag ggtggcccc tgtgccagct caggtgactg gagcccagg gcctgacagg 2220
ttcccagcag ctgggcgggc cagccttgca ctgtgggggc tggctcctgag tcttggcctg 2280
cctcccagcc ctgccaggg gctgcggcct aggggtccac gggaagcctc cgtcaggaga 2340
gacgcagccc tgggggccc ctggtgctgc aaggaagggg gggaagcttg ctggcttctg 2400
ttgcgccact gagacggcag agaccccagg atcccagagc tcccaggat ccctcccaga 2460
tcctctgctg actccatat gaggcctcac acccagagg tagggcagca gatcttctt 2520
ataactattt attgttcgaa tcacttttag gatgtaactt tataaataaa catgagcgct 2580
gaaaaaaaaa agggg 2594
```

&lt;210&gt; 34

&lt;211&gt; 481

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2024210



```

<400> 34
tgaggtgcga cacacataat tgtcccaatt tttaagattg atggggagca tgaagcattt 60
ttttaatgtg ttggcaggcc ccattaaatg cataaactgc ataggactca tgtgggtctga 120
atgtatttta gggctttctg ggaattgtct tgacagagaa cctcagctgg acaaagcagc 180
cttgatctga gtgagctaac tgacacaatg aaactgtcag gcatgtttct gctcctctct 240
ctggctcttt tctgcttttt aacagggtgtc ttcagtcagg gaggacagggt tgactgtggg 300
gagttccagg accccaagggt ctactgcact cgggaatcta acccacactg tggctctgat 360
ggccagacat atggcaataa atgtgccttc tgtaaggcca tagtgaaaag tgggtggaaag 420
attagcctaa agcatcctgg aaaatgctga gttaaagcca atgtttcttg gtgacttgcc 480
a 481

```

```

<210> 35
<211> 3080
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte Clone No: 2523109

```

```

<400> 35
cgcagcgcg ctttcaggcc aacatggccg tgctgctgct gctgctccgt gccctccgcc 60
gggggtccagg cccgggtcct cggcgcgtgt gggggccagg cccggcctgg agtccagggt 120
tccccgccag gcccgggagg gggcggccgt acatggccag caggcctccg ggggacctcg 180
ccgaggtgg aggccgagct ctgcagagct tacaattgag actgctaacc cctacctttg 240
aagggatcaa cggattgttg ttgaaacaac atttagttca gaatccagtc agactctggc 300
aacttttagg tggtaactttc tattttaaca cctcaagggt gaagcagaag aataaggaga 360
aggataagtc gaaggggaag gcgcctgaag aggacgaaga ggagaggaga cgccgtgagc 420
gggacgacca gatgtaccga gagcggctgc ccaccttgct ggtcatcgcg gttgtcatga 480
gcctcctgaa tgctctcagc accagcggag gcagcatttc ctggaacgac tttgtccacg 540
agatgctggc caagggcgag gtgcagcgcg tccagggtgg gcctgagagc gacgtgggtg 600
aagtctacct gcacctgga gccgtgggtg ttggggcgcc tcggctagcc ttgatgtacc 660
gaatgcagggt tgcaaatatt gacaagtttg aagagaagct tcgagcagct gaagatgagc 720
tgaatatcga ggccaaggac aggatcccg tttcctacaa gcgaacagga ttctttggaa 780
atgccctgta ctctgtgggg atgacggcag tgggcctggc catcctgtgg tatgttttcc 840
gtctggccgg gatgactgga agggaagggt gattcagtg ttttaatcag cttaaaatgg 900
ctcgtttcac cattgtggat gggaagatgg ggaaaggagt cagcttcaaa gacgtggcag 960
gaatgcacga agccaaactg gaagtcgcgc agtttgtgga ttatctgaag agcccaaaac 1020
gcttctcca gcttggcgcc aaggteccaa agggcgcact gctgctcggc cccccgggt 1080
gtgggaagac gctgtggcc aaggcggtgg ccacggaggc tcagggtgcc ttctggcgga 1140
tggcgggccc agagttcgtg gaggtcattg gaggcctcgg cgctgcccggt gtgcgagacc 1200
tctttaagga agcccgagcc cgggccccct gcacgtgcta catcgatgag atcgacgagg 1260
tgggcaagaa gcgctccacc accatgtccg gcttctccaa cacggaggag gagcagagc 1320
tcaaccagct tctggtagaa atggatggaa tgggtaccac agaccatgtc atcgtcctgg 1380
cgtccacgaa ccgagctgac attttgagc gtgctctgat gaggccaggc cgactggacc 1440
ggcacgtctt cattgatctc cccacgctgc aggagaggcg ggagattttt gagcagcacc 1500
tgaagagcct gaagctgacc cagtccagca ccttttactc ccagcgtctg gcagagctga 1560
caccaggatt cagtggggct gacatcgcca acatctgcaa tgaggctgcg ctgcacgagg 1620
cgcgggaggg acacacttcc gtgcacactc tcaacttcga gtacgccgtg gagcgcgctc 1680
tcgcagggac tgccaaaaag agcaagatcc tgtccaagga agaacagaaa gtggttgctg 1740
ttcatgagtc gggccacgcc ttggtgggct ggtgctgga gcacacggag gccgtgatga 1800
aggctctcat aaccctcgg acaaacgcgc cctgggctt tgctcagatg ctccccagag 1860
accagcacct cttaccaag gagcagctgt ttgagcggat gtgcatggcc ttggaggagc 1920
gggcctcgga agcactgtcc ttcaacgagg tcacttctgg ggcacaggac gacctgagga 1980
aggtcacccg catcgcttac tccatggtga agcagtttgg gatggcacct ggcacgggc 2040
ccatctcctt ccttgaggcg caggaggggc tcatgggcat cgggcggcgc ccttcagcc 2100
aaggcctgca gcagatgat gaccatgaag caagactgct ggtggccaag gcctacagac 2160

```

```

acaccgagaa ggtgctgcag gacaacctgg acaagttgca ggcgctggca aacgcccttc 2220
tggaaaagga agtgataaac tatgaggaca ttgaggtctt cattggcccg ccgccccatg 2280
ggccgaagaa aatgatcgca ccgcagaggt ggatcgacgc ccagagggag aaacaggact 2340
tgggcgagga ggagaccgaa gagaccagc agcctccact tggaggcgaa gagccgactt 2400
ggcccaagta gttgggaggt gttggctgca cgtgcgggtg gtccgggaag tgagggtctca 2460
ctcagccacc ctgagttgct tttcagctga ggtttgact tcctctcgcg gccctcagta 2520
gtccctgcac agtgacttct gagatctgtt gattgatgac ccttttcagtg attttaagtt 2580
tctctgcaga aactactgac ggagtcctgt gtttgtgagt cgtttccctt atggggaagg 2640
ttatcagtgc ttcccagtg agcatggaac acttcgagtt cccagggtta tagacagtcg 2700
ttcccagtg ggctgaggcc acccagaggc agcagagcat tcagactcca aacagacccc 2760
tgttcatgcc gacgcttgca cgaccgccc agttcctgtg gtcctctcgg aatgctaagg 2820
ggatcggaaca tgaaaggacc ctgtgagccg attgtcctat ctccagcggc cctgtcatcc 2880
agctcactca tcaatggggc cacacagtca ggccaggca ctgggctccg gaggactcac 2940
cactgcccc tgctgccatg tggactggtg caagttgagg acttcttgct ggtctagtca 3000
cgcattgcagt gttggggatg ccttggtttt tactgctctg agaattgttg agatacttta 3060
ctaataaact gtgtagttgg                                     3080

```

&lt;210&gt; 36

&lt;211&gt; 1154

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2588566

&lt;400&gt; 36

```

gaactatggt gtggttgac agacacggag aaaatcagcg gagttcttgc tccgaatctt 60
cctgtaaatg ccagacagtg acaggcacct gagcagccat ttcaacctca gaatgaaggg 120
aagcccttca gaacatggct cccaacaaag ctttttcaac agatatgctc agcagaggct 180
ggacattgat gccacccagc ttcaggccct tctcaaccag gagcttctaa caggacctcc 240
aggggacatg ttctccttag atgagtgccg cagcttggtg gctctgatgg aactgaaagt 300
gaatgggagg ctagaccaag aggagtttgc gcgactgtgg aagcgcttg ttcactacca 360
gcatgttttc cagaaggttc agacaagccc tggagtcctc ctgagctcgg acttgtggaa 420
ggccatagag aatacagact tctcagagg gatcttcac agccgtgagc tgetgcatct 480
ggtgaccctc aggtacagcg acagcgtcgg cagggtcagc ttccccagcc tgggtctgctt 540
cctgatgcgg cttgaagcca tggcaaagac ctccgcaac ctctctaagg atggaaaagg 600
actctacctg acagaaatgg agtggatgag cctggctcatg tacaactgaa gcaaagagga 660
aagcagaccc atggctcagg acaagctccc agtgatcact caagaatctg gctctcattc 720
taagaggctg tgetgcccag tatggtggtt gtgataaatc taaaccagcc ctgcatgaaa 780
cagagtccaa gctgtctccc aacagcctgg gttcggctct tggctggccc aggccagtt 840
aagcctgtgg ccaccaagca gctcatctga gcactttggg atgtattcag cctacgttgc 900
cctggaaaag gaagcaggag atgtctccct gtgggaaagg agaagagaag ttgtctctga 960
gtcccctgtc accagttgga ttcatctctt ggaagagcca gaatgagcca ctttgaccac 1020
cctcgggtgc tatgggtgac acaagagctg tccactgggt gtttgagaaa taattacact 1080
atcttatgtc tggatcctga tgatttcaca gctaaatggc aaaaataaaa catgtttccc 1140
ataaaaaaaaaaaaaa                                     1154

```

&lt;210&gt; 37

&lt;211&gt; 2827

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; 2811

&lt;223&gt; a or g or c or t, unknown, or other

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2740570

&lt;400&gt; 37

```
catttcggggg gtattctcag caggatgctc aagaattcct tcgatgttta atggatttgc 60
ttcatgaaga attgaaagag caagtcatgg aagtagaaga agatccgcaa accataacca 120
ctgaggagac aatggaagaa gacaagagcc agtcggatgt agattttcag tcttgtgaat 180
cttgtagcaa cagtgataga gcagaaaatg aaaatggctc tagatgcttt tctgaagata 240
ataatgaaac aacaatgtta attcaggatg atgaaaacaa ttcagaaatg tcaaaggatt 300
ggcaaaaaga gaagatgtgc aataagatta ataaagtaaa ttctgaaggc gaatttgata 360
aagatagaga ctctatatct gaaacagtcg acttaaacaa ccaggaaact gtcaaagtgc 420
aaatacacag cagagcttca gaatatatca ctgatgtcca ttcgaatgac ctgtctacac 480
cacagatcct tccatcaaat gaagggtgta atccacgttt atcggcaagc cctcctaaat 540
caggcaattt gtggccagga ttggcaccac cacacaaaaa agctcagtc gcctctccaa 600
agagaaaaaa acagcacaaag aaatacacag agtgtatttc agacatatt gatggaacaa 660
tcattagttc agtgcagtg ctgacttggt acaggggtgc tgtaaccctc gagacctttc 720
aagatctgtc cttgccaatt cctggcaagg aagaccttgc taagctgcat tcatcaagtc 780
atccaacttc tatagtcaaa gcaggatcat gtggcgaagc atatgctcca caagggtgga 840
tagctttttt catggaatat gtgaagaggt ttgttgtctc atgtgtccct agctggtttt 900
ggggtccagt agtaaccttg caagattgtc ttgctgcctt ctttgccaga gatgaactaa 960
aagggtgacaa tatgtacagt tgtgaaaaat gcaaaaagct gagaaatgga gtgaagtttt 1020
gtaaaagtaca aaactttcct gagattttgt gcattccact taaaagattc agacatgaac 1080
taatgttttc caccaaaatc agtacctatg ttctatttcc gctagaaggc ttggatcttc 1140
agccatttct tgctaaggat agtccagctc aaattgtgac atatgatctt ctgtcagtca 1200
tttgccatca tggaaactgca agtagtgagc actatatagc ctactgccga aacaatctaa 1260
ataatctctg gtatgaattt gatgatcaga gtgtcactga agtttcagaa tctactgtac 1320
aaaatgcaga agcttacgtt cttttctata ggaagagcag cgaagaggca caaaaagaga 1380
ggagaaggat atcaaattta ttgaacataa ttgaaacaa cctccttcag ttttatattt 1440
ctcgacagtg gcttaataaa tttaagacct ttgccgaacc tggccctatt tcaaataatg 1500
actttctttg tattcatgga ggtgttcctc caagaaaagc tggttatatt gaagacctgg 1560
ttttgatgct gcctcagaac atttgggata acctatatag cagggtatggg ggaggaccag 1620
ctgtcaacca tctgtacatt tgtcatactt gccaaattga ggcggagaaa attgaaaaaa 1680
gaagaaaaaac tgaattggaa atttttattc ggcttaacag agcgttccaa aaagaggact 1740
ctccagctac tttttattgc atcagtatgc agtgggttag agaatgggaa agttttgtga 1800
agggtaaaga tggagatcct ccaggtccta ttgacaatac taagattgca gtcactaaat 1860
gtggtaatgt gatgcttagg caaggagcag attctggcca gatttctgaa gaaacatgga 1920
attttctgca gtctatttat ggtggagggg ctgaagttat cctgcgacct ccggttgttc 1980
atgttgatcc agatatactt caagcagaag aaaaaattga agtagaaact cggctcttgt 2040
aatttttagg atgtagagag ttctaataag gaatcatttt catgtgccct gacatgtaca 2100
catgcgaaaa cattcctaaa agcgtgttta tttgctttat ttttttcat catttatccc 2160
atttatttct tcttagtggg cattatggaa gaatatatta aaatgtgtaa tataccacag 2220
gttgggtatat ttagttttta atacttacca taaagtcttt cagtgttaatt tttttttgag 2280
acagagtctt gctttgtcac ccaggctgga gtgctgtggt gttacctcag ctactgcag 2340
cctccacctc ctgggttcaa gcgattctcc tgcctcagcc tctcgagtag ctgggattac 2400
aggcacctgc caccatgccc ggctaatttt tgtattttag tagagatggg gtttcaccat 2460
gttggccagg ctagtctcaa actcctgacc tcagggtgac caccacactg ggcctcccaa 2520
agtgtcggga ttacaggtgt gagccacagc gcctggctcg tgttctaagg aaatagctac 2580
ctaaggatc ttattaaaac aaatgaacaa aaagtttccc aaactgtgtt ctaaggaaat 2640
agctacctat cttattaaaa aaaaaacaaa aacaaagttt tatgctctaa taagtttggg 2700
aaatgctagg ttatacaaag ctaaacatga ttctttccag tgggacatct gagagtcttt 2760
aattagctaa cagtgcattg tgattctgca aaggagttca ataattcacc ngtacgcgtg 2820
gtcgacc
```

&lt;210&gt; 38

&lt;211&gt; 2987

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte Clone No: 2820384

&lt;400&gt; 38

```
tgctcaccta cgcagctcct ctgcgccctg cagccccgtc caccaccacg agggccatgc 60
caagctgtct agcagccccc ctctgtgcaag ccccgtaggg gtggcaccgt cgtacgtgct 120
caagaaaagg caggtattga gtgctggcag ccggaggggt aaggagcagc gctaccgcag 180
cgtcatctca gacatctttg acggctccat tcttagcctc gtgcagtgtc tcacctgtga 240
ccgggtatcc accacagtgg aaacgttcca ggacttatca ctgcccattc ctggaaagga 300
ggacctggcc aagctccatt cagccatcta ccagaatgtg ccggccaagc caggcgctg 360
tggggacagc tatgccgccc agggctggct ggcccttcat gtggagtaca tccgacgggt 420
tgtggtatcc tgtacacca gctggttttg ggggcctgtc gtcaccctgg aagactgcct 480
tgctgccttc tttgccgctg atgagttaaa gggtgacaac atgtacagct gtgagcgggt 540
taagaagctg cggaacggag tgaagtactg caaagtcctg cggttgcccc agatcctgtg 600
cattcaccta aagcgctttc ggcacgaggt gatgtactca ttcaagatca acagccacgt 660
ctccttcccc ctcgaggggc tcgacctgcg ccccttccct gccaaaggag gcacatccca 720
gatcaccacc tacgacctcc tctcggatcat ctgccaccac ggcacggcag gcagtgggca 780
ctacatcgcc tactgccaga acgtgatcaa tgggcagtgg tacgagtttg atgaccagta 840
cgtcacagaa gtccacgaga cgggtgtgca gaacgccgag ggctacgtac tcttctacag 900
gaagagcagc gaggagggca tgcgggagcg acagcaggtg gtgtccctgg ccgccatgcg 960
ggagcccagc ctgctgcggt tctacgtgtc ccgcgagtggt ctcaacaagt tcaacacctt 1020
cgcagagcca ggccccatca ccaaccagac ctctcctctg tcccacggag gcatccccgc 1080
ccacaaatac cactacatcg acgacctggt ggtcatcctg cccagaacg tctgggagca 1140
cctgtacaac agattcgggg gtggccccgc cgtgaaccac ctgtacgtgt gctccatctg 1200
ccaggtggag atcgaggcac tggccaagcg caggaggatc gagatcgaca ccttcatcaa 1260
gttgaacaag gccttccagg ccgaggagtc gccgggctgc atctactgca tcagcatgca 1320
gtggttccgg gagtgggagg cgttcgtcaa ggggaaggac aacgagcccc ccgggccccat 1380
tgacaacagc aggattgcac aggtcaaagg aagcggccat gtccagctga agcagggagc 1440
tgactacggg cagatttcgg aggagacctg gacctacctg aacagcctgt atggaggtgg 1500
ccccgagatt gccatccgcc agagtgtggc gcagcgctgg gccagagaa cctgcacggg 1560
gagcagaaga tcgaagccga gacgggggcc gtgtgatctg ctgggctagt cccccatgt 1620
gccccacccc gcggaaggcg tgtttgtgcc cagaagagag gccgggctgc tgcagaaccc 1680
cgccgtgtaa agaggcagaa aagtttggtt ggtttgcagt aacgtgcaa ctagaaaata 1740
tatgcacttc aggtctgttg aaacgaccaa gactctgtga cgttaatttg ggtctttgtc 1800
ctggcagtgc ctctgccagt cactgtcatc gttgtgtccc ccacaactgt cctcttcta 1860
gctcgcccca gctttgtccc tggagccga tgcataccct gtcagacaga ggctgcggcc 1920
tgggccagag tcagggagta gctgctgctt cacggcgtct cactgtgctg attggcccg 1980
agccccgaag actcggaggg agctgctcag ggccggtag cgcaccagaa gccctggcca 2040
gtgaggagct cacaggtcct ccctgggtgt cccgcgcac ctctgcatct cctgggctgc 2100
accaggaagg ctctgaagtc ccgggtgct ctcagcactt ctctgcaga ctgaagactc 2160
tggaactcatt gctgattgga acaccaggag gaggttggtt ttctgccagt gggggatgtt 2220
tctggaggca gctggctccc cacaccgcgt cctgctgagc ctgccccctg gattggctgt 2280
aatttgcctc gaagttcagc agttcatctt catgggaaat ttgctgagcc cccaccagg 2340
aaccggatga tgaacaggg atacctcaca gacttttcc ctctctccca gcagcagcag tgcagagccc 2400
cgagctgatg ctaaagaaga cagactttcc ctctctccca gcagcagcag tgcagagccc 2460
gcctggaggg atgtgggggc tgtgcagggt gcagcgctca ggtggatcct ggggaagcagc 2520
ctctggatgc tgagtggagg gagccactga gcacagcaag gcaccaaagc cctggagaa 2580
accgccaggg cgaggtgcga ccatcatcag gatcaaagca gacggggcgt ggggtgggaa 2640
ggggctctgg gaccagaccc cccacactac tgcgtctttg tttctatcag tctttgtaga 2700
agcaggtggt ggtggaaatt ccagcaggtg ggtcccgtag aggccttag gcctcacttt 2760
tcggatcttc tgtcccagat cctgctccct cctgctgag cctgggggtt cctggcatt 2820
ggccccagcc ttctgaaagc cggcgctgca gccagaggcc gcacgctgca ctgtcgcgac 2880
gcagagaggc ttctgtgcag gctgggatcg ggccccatgt ctgtgctgtc tagtttgtgt 2940
tcaaatgtc agaataaaca cagaataaat gttaaaaaaa aaaaaaa 2987
```

<210> 39  
 <211> 1215  
 <212> DNA  
 <213> Homo sapiens  
 <220>  
 <221> misc\_feature  
 <223> Incyte Clone No: 2990692

<400> 39  
 agagcacctt agtaggccgg attcgggtca gatgagtatg cataaggcaa tgctaattggc 60  
 tcaagcaatg aggggggtca ctctaggagg acaagttaga acatttgga aaaaatgtta 120  
 taattgtgtt caaatcggtc atctgaaaag gagttgcccc gtcttaaata aacagaatat 180  
 aataaatcaa gctattacag caaaaaataa aaagccatct ggctgtgtc caaaatgttg 240  
 aaaaggaaaa cattgggcca atcaatgtca ttctaaattt gataaagatg ggcaaccatt 300  
 gtcgggaaac aggaagaggg gccagcctca ggcccccaa caaactgggg cattcccagt 360  
 tcaactgttt gttcctcagg gttttcaagg acaacaaccc ctacagaaaa taccaccact 420  
 tcagggagtc agccaattac aacaatccaa cagctgtccc gcgccacagc aggcagcgcc 480  
 acagtagatt tatgttccac ccaaatgggtc tctttactcc ctggagagcc cccacaaaag 540  
 attcctagag ggggtatatg cccgctgcca gaaggagggg taggccttat tttagggaga 600  
 tcaagtctaa atttgaaggg agtccaaatt catactgggg taatttattc agattataaa 660  
 gggggaattc agttagtgat cagctccact gttccctgga gtgccaatcc aggtgataga 720  
 attgctcaat tactgctttt gccttatgtt aaaattgggg aaaacaaaac ggaaagaaca 780  
 ggagggtttg gaagtaccaa ccctgcagga aaagccactt attgggctaa tcagggtctca 840  
 gaggatagac ccgtgtgtac agtcaactatt ccagggaag agtttgaagg attagtggat 900  
 acccaggctg atgtttctat catcggcata ggcaccgcct cagaagtgt tcaaagtgcc 960  
 atgattttac attgtctagg atctgataat caagaaagta cggttcagcc tatgatcact 1020  
 tctattccaa tcaatttatg gggccgagac ttgttacaac aatggcatgc agagattact 1080  
 atcccagcct ccctatacag cccaggaat caaaaaatca tgactaaaat gggatagctc 1140  
 cctaaaaagg gactaggaaa gaatgaagat ggcattaaag tcccaactga ggctgaaaaa 1200  
 aatcaaaaaa aaaaa 1215

<210> 40  
 <211> 1037  
 <212> DNA  
 <213> Homo sapiens  
 <220>  
 <221> misc\_feature  
 <223> Incyte Clone No: 4590384

<400> 40  
 gccatggggc tggggttgag gggctgggga cgtcctctgc tgactgtggc caccgcccctg 60  
 atgctgcccc tgaagcccc cgcaggctcc tggggggccc agatcatcgg gggccacgag 120  
 gtgaccccc actccaggcc ctacatggca tccgtgcgct tcgggggcca acatcactgc 180  
 ggaggcttcc tgetgcgagc ccgctgggtg gtctcgcccg cccactgctt cagccacaga 240  
 gacctccgca ctggcctggt ggtgctgggc gccacgctcc tgagtactgc ggagccacc 300  
 cagcaggtgt ttggcatcga tgcctcacc acgcaccctg actaccaccc catgaccac 360  
 gccaacgaca tctgcctgct gcggctgaac ggtctgctg tccctgggccc tgcagtgggg 420  
 ctgctgaggg tgccaggag aagggccagg cccccacag cggggacag gtgccgggtg 480  
 gctggctggg gcttcgtgtc tgactttgag gagctgcgcg ctggactgat ggaggccaag 540  
 gtccgagtgc tggaccgga cgtctggaac agtctctgga agggccacct gacacttacc 600  
 atgctctgca cccgcagtgg ggacagccac agacggggct tctgctcggc cgactccgga 660  
 gggcccctgg tgtgcaggaa ccgggctcac ggcctcgttt ccttctcggg cctctgggtg 720  
 ggcgacccca agacccccga cgtgtacacg cagggtgtccg cctttgtggc ctggatctgg 780  
 gacgtggttc ggcggagcag tccccagccc ggccccctgc ctgggaccac caggcccca 840  
 ggagaagccg cctgagccac aaccttgccg catgcaaatg agatggccgc tccaggcctg 900

```
gaatgttccg tggctgggcc ccacgggaag cctgatgttc agggttgggg tgggacgggc 960
agcgggtgggg cacaccatt ccacatgcaa agggcagaag caaaccagt aaaatgttaa 1020
ctgacgaaaa aaaaaaa                                1037
```



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> : <b>C12N 15/57, 9/48, C12Q 1/68, C12N 5/10, C07K 16/40, A61K 38/48, 38/55</b>		<b>A3</b>	(11) International Publication Number: <b>WO 00/09709</b>
			(43) International Publication Date: 24 February 2000 (24.02.00)
(21) International Application Number: PCT/US99/17818 (22) International Filing Date: 6 August 1999 (06.08.99) (30) Priority Data: 60/096,114           10 August 1998 (10.08.98)      US 60/119,768           11 February 1999 (11.02.99)      US (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications US                                   60/096,114 (CIP) Filed on                           10 August 1998 (10.08.98) US                                   60/119,768 (CIP) Filed on                           11 February 1999 (11.02.99) (71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577		(US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). REDDY, Roopa [IN/US]; 1233 W. McKinley Drive, Sunnyvale, CA 94086 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). AU-YOUNG, Janice [US/US]; 1419 Kains Avenue, Berkeley, CA 94702 (US). SHIH, Leo, L. [US/US]; Apartment B., 1081 Tanland Drive, Palo Alto, CA 94303 (US). LU, Dyung, Aina, M. [US/US]; 55 Park Belmont Place, San Jose, CA 95136 (US). (74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published With international search report. (88) Date of publication of the international search report: 8 September 2000 (08.09.00)	
(54) Title: PROTEASES AND ASSOCIATED PROTEINS			
(57) Abstract			
The invention provides human proteases and associated proteins (PPRG) and polynucleotides which identify and encode PPRG. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of PPRG.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		



# INTERNATIONAL SEARCH REPORT

Inte. onal Application No  
PCT/US 99/17818

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/57 C12N9/48 C12Q1/68 C12N5/10 C07K16/40  
A61K38/48 A61K38/55

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BERNOT ET AL.: "Metalloproteinase" EMBL SEQUENCE DATABASE, 1 June 1998 (1998-06-01), XP002129490 HEIDELBERG DE	1-6,9-11
Y	Ac 043923 the whole document	1-16,19, 20
X	-& BERNOT ET AL.: "Homo sapiens mRNA for metalloproteinase" EMBL SEQUENCE DATABASE, 21 January 1998 (1998-01-21), XP002129491 HEIDELBERG DE	1-6,9-11
Y	Ac AJ003144 the whole document	1-16,19, 20
X	-& BERNOT ET AL.: "A transcriptional map of the FMF region" GENOMICS, vol. 50, 1998, pages 147-160, XP002090815	1-6,9-11
Y	page 154, left-hand column	1-16,19, 20
-/-		



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

\* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*Z\* document member of the same patent family

Date of the actual completion of the international search

3 February 2000

Date of mailing of the international search report

15. 5. 00

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

CEDER O.

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/17818

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p style="text-align: center;">---</p> <p>WO 96 04387 A (ROUSSEL UCLAF ;DIU ANITA (FR); FAUCHEU CHI (FR); HERCEND THIERRY ()) 15 February 1996 (1996-02-15) page 2, line 31 -page 3, line 15 page 8, line 39 -page 11, line 5 abstract; claims 1-6,11-17</p> <p style="text-align: center;">-----</p>	<p>1-16,19, 20</p>

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 99/ 17818

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
see FURTHER INFORMATION PCT/ISA/210
2. ☒ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
see FURTHER INFORMATION PCT/ISA/210
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-16, 19-20 ALL PARTIAL

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

## Continuation of Box I.1

Although claims 19 and 20 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

-----

## Continuation of Box I.1

Rule 39.1(iv) PCT - Method for treatment of the human or animal body by therapy

-----

## Continuation of Box I.2

Claims Nos.: 17 18

Claims 17 and 18, relating to purified agonists/antagonists of the polypeptide of claim 1, could not be searched as its subject-matter was insufficiently disclosed.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

**FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210**

**1. Claims: Invention 1: 1-16 19-20 all partial**

A substantially purified polypeptide and an isolated and purified polynucleotide encoding it and uses of them, where the polypeptide sequence is SEQ ID NO 1 and the the polynucleotide sequence is SEQ ID NO 21.

**2. Claims: Invention 2-20: 1-16 19-20 all partial**

Idem as subject 1 but limited to each of the polypeptide sequences as in SEQ ID NOS 2-20 and corresponding polynucleotide sequences SEQ ID NOS 21-40, where invention 2 is limited to SEQ ID NOS 2 and 22, invention 3 is limited to SEQ ID NOS 3 and 23, ....., invention 20 is limited to SEQ ID NOS 20 and 40.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 99/17818

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9604387 A	15-02-1996	FR 2723378 A	09-02-1996
		AU 704426 B	22-04-1999
		AU 3118095 A	04-03-1996
		CA 2196339 A	15-02-1996
		CN 1158639 A	03-09-1997
		EP 0774004 A	21-05-1997
		HU 76971 A	28-01-1998
		JP 10503652 T	07-04-1998
		US 6020477 A	01-02-2000